DOCUMENT RESUME

ED 388 672 TM 023 189

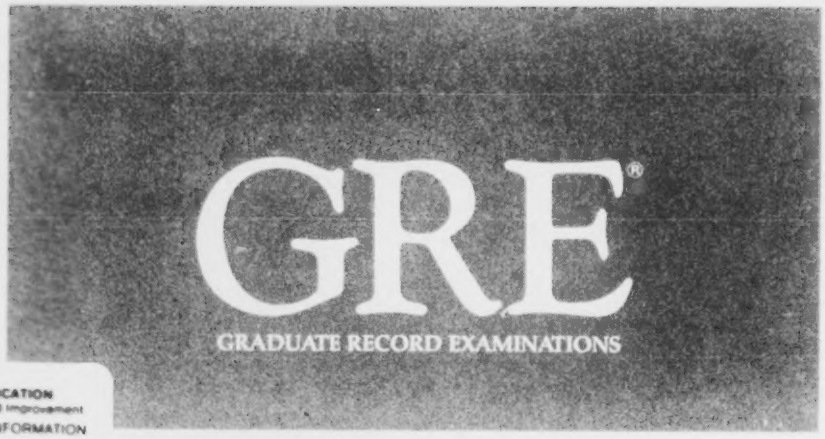
AUTHOR	Zwick, Rebecca; Braun, Henry I.
TITLE	Methods for Analyzing the Attainment of Graduate School Milestones: A Case Study. GRE Board Professional Report No. 86-3P.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	ETS-RR-88-30
PUB DATE	Jun 88
NOTE	113p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Doctoral Programs; *Educational Attainment; Enrollment Trends; Ethnic Groups; *Foreign Students; Grade Point Average; *Graduate Students; Graduate Study; *Graduation; Higher Education; Longitudinal Studies; Predictor Variables; Sex Differences; *Test Results
IDENTIFIERS	Graduate Record Examinations

ABSTRACT

Using a data base compiled by the Graduate School of Northwestern University, a longitudinal study of the graduate school careers of 2,211 students in 14 programs was conducted. Among the most prominent findings was the increase in the enrollment of foreign students. The patterns of attainment of graduate school milestones, such as Ph. D. candidacy and graduation were examined for each graduate program and for gender and ethnic groups. There was substantial variation across programs and, to a lesser degree, across demographic groups. Graduation rates for foreign students were higher than those for U.S. citizens. The association between the attainment of milestones and measures of academic potential, such as undergraduate grade point average (UGPA) and Graduate Record Examination (GRE) scores, was also investigated. The likelihood of attaining candidacy or of completing a doctorate was found to bear little relation to UGPA and GRE scores. This finding is probably a result of the use of UGPA and GRE in the selection of students into graduate programs. Appendix A presents 28 tables of ethnic and gender composition. Appendix B discusses survival analysis. Appendix C contains an empirical Bayes strategy for logistic regression. (Contains 20 figures, 7 tables in the text, and 22 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 388 672



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

METHODS FOR ANALYZING THE ATTAINMENT OF GRADUATE SCHOOL MILESTONES: A CASE STUDY

Rebecca Zwick
and
Henry I. Braun

GRE Board Professional Report No. 86-3P
ETS Research Report 88-30

June 1988

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

TM023189

BEST COPY AVAILABLE

Methods for Analyzing the Attainment of Graduate School Milestones: A Case Study

Rebecca Zwick
Henry I. Braun

GRE Board Report No. 86-3P

June 1988

Educational Testing Service, Princeton N.J. 08541

Copyright © 1988 by Educational Testing Service. All rights reserved.

Acknowledgments

We wish to thank Ka-ling Chan, who performed data analyses, Dorothy Thayer, who wrote and executed the survival analysis program, and Ira Sample, who assisted in preparing graphs. We are also indebted to Clarence Ver Steeg, David Cohen, Sonya Bowarchuk, and Robin Gillies of Northwestern University for providing us with the data and for offering ongoing consultation. This study was funded by the Graduate Record Examinations Board.

Abstract

Using a data base compiled by the Graduate School of Northwestern University, a longitudinal study of the graduate school careers of students in fourteen programs was conducted. Among the most prominent findings was the increase in the enrollment of foreign students. The patterns of attainment of graduate school milestones, such as Ph.D. candidacy and graduation, were examined for each graduate program and for gender and ethnic groups. There was substantial variation across programs and, to a lesser degree, across demographic groups. Graduation rates for foreign students were higher than those for U.S. citizens. The association between the attainment of milestones and measures of academic potential, such as undergraduate grade-point average (UGPA) and Graduate Record Examination (GRE) scores was also investigated. The likelihood of attaining candidacy or of completing a doctorate was found to bear little relation to UGPA and GRE scores. This finding is probably a result of the use of UGPA and GRE in the selection of students into graduate programs.

To many observers, graduate education in the United States is at a critical juncture. Recent studies show that American students represent a decreasing percentage of students enrolled in U.S. graduate schools. Another troublesome trend is the decrease in Black enrollment during the last decade (Brown, 1987; Trent & Copeland, 1987). There has been a growing concern that talented undergraduates may be choosing to go to professional schools or turning immediately to the world of work upon graduation (see Hartnett, 1987). Undoubtedly there is a complex web of causes underlying these patterns, among them the financial burdens education imposes, perceived job opportunities and the vicissitudes of fashion. Graduate school deans are now faced with the challenge of analyzing these trends and developing appropriate policies.

It is important, therefore, to determine what happens to those individuals who actually enroll in graduate school. At what pace do these students reach milestones in their graduate careers, such as advancement to candidacy and attainment of the Ph.D. degree? What attributes differentiate students who complete the doctorate from those who do not? How do the patterns of achievement differ across academic programs? Answers to these questions about pathways through graduate school can provide information that will be useful to graduate school policymakers in allocating resources and improving educational practices.

To investigate these issues, we used a unique data base from the Graduate School of Northwestern University that can support longitudinal cohort analyses of graduate school careers. This date base contains the records of applicants to the Graduate School over a period of nearly fifteen years.

Our research focused on fourteen graduate programs: Chemical Engineering, Computer Science, Chemistry, Mathematics, Physics, Counseling Psychology, Clinical Psychology, Sociology, Theatre, English, History, Political Science, Economics, and Philosophy. Chemical Engineering and Computer Science are part of the Technological Institute, Counseling Psychology is part of the School of Education, Clinical Psychology is part of Clinical Medicine, and Theatre is part of the School of Speech. The remaining programs come under the rubric of Arts and Sciences. These programs were selected because they are of general interest, their sample sizes are adequate, and they are thought to be relatively free of major administrative shifts during the time period in question. Only students who stated at entry that they were seeking a Ph.D. were included in the study.

Research Questions

Our research questions fell into two broad categories:

1. How do the patterns of attainment of graduate school milestones, such as Ph.D. candidacy and graduation, differ across academic disciplines and across demographic subgroups?

2. What is the association between students' attainment of milestones in their graduate careers and measures of their academic potential, such as undergraduate grade-point average (UGPA) and Graduate Record Examination (GRE) scores?

Because our study focuses on a single school only, our findings cannot be assumed to have broad applicability. Rather, our research illustrates the kinds of analyses that may be useful in addressing policy decisions about enrollment, retention, and academic policy in graduate schools.

Data Analysis

Introduction

Our data analyses are of three basic types. First, descriptive analyses were conducted, showing the numbers of students entering each of the 14 graduate programs, the proportions of women, minorities, and foreign students, and the candidacy and graduation rates for various groups. The second category of analyses involves the examination of patterns of attainment of graduate school milestones for each of the graduate programs. The final phase of analysis involves investigation of the association between attainment of milestones and potential explanatory variables. Although the data base included students who entered between 1972 and 1986, some analyses were based on only a subset of these students. Details are given in the following sections.

Descriptive Analyses

Tables 1 and 2 provide information about the demographic makeup of Northwestern students in the 14 selected graduate programs for the entering classes of 1975 through 1986. (Students who entered during the years 1972-1974 were not included because accurate demographic information was unavailable.) Entry years have been grouped into four sets of three years. Table 1 provides ethnic information and Table 2 gives the proportions of male and female students for the 14 graduate programs combined. Tables A1-A28 (Appendix A) provide corresponding information for each of the 14 programs.

In the Northwestern data base, students are assigned to one of the following categories, usually on the basis of self-report:

1 American Indian
2 Black
3 Oriental
4 Hispanic
5 Foreign
6 White
7 Mexican-American/Chicano
8 Puerto Rican

Note that no ethnic information is available for foreign students. In the

Table 1

Ethnic Composition for 1975-1986 Entrants[a]:
All 14 Graduate Programs Combined

		WHITE	BLACK	ASIAN	HISPANIC	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	395	22	5	19	101	113	655
	ROW%	60.31%	3.36%	0.76%	2.90%	15.42%	17.25%	100.00%
	COL%	27.19%	33.85%	10.42%	54.29%	20.78%	91.13%	29.62%
REGYR 1978-80	N	398	19	7	6	104	9	543
	ROW%	73.30%	3.50%	1.29%	1.10%	19.15%	1.66%	100.00%
	COL%	27.39%	29.23%	14.58%	17.14%	21.40%	7.26%	24.56%
REGYR 1981-83	N	305	15	18	4	102	2	446
	ROW%	68.39%	3.36%	4.04%	0.90%	22.87%	0.45%	100.00%
	COL%	20.99%	23.08%	37.50%	11.43%	20.99%	1.61%	20.17%
REGYR 1984-86	N	355	9	18	6	179	0	567
	ROW%	62.61%	1.59%	3.17%	1.06%	31.57%	0.00%	100.00%
	COL%	24.43%	13.85%	37.50%	17.14%	36.83%	0.00%	25.64%
TOTAL	N	1453	65	48	35	486	124	2211
	ROW%	65.72%	2.94%	2.17%	1.58%	21.98%	5.61%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table 2

Proportions of Men and Women for 1975-1986 Entrants:
All 14 Graduate Programs Combined

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	421	218	16	655
	ROW%	64.27%	33.28%	2.44%	100.00%
	COL%	28.37%	30.70%	94.12%	29.62%
REGYR 1978-80	N	373	169	1	543
	ROW%	68.69%	31.12%	0.18%	100.00%
	COL%	25.13%	23.80%	5.88%	24.56%
REGYR 1981-83	N	313	133	0	446
	ROW%	70.18%	29.82%	0.00%	100.00%
	COL%	21.09%	18.73%	0.00%	20.17%
REGYR 1984-86	N	377	190	0	567
	ROW%	66.49%	33.51%	0.00%	100.00%
	COL%	25.40%	26.76%	0.00%	25.64%
TOTAL	N	1484	710	17	2211
	ROW%	67.12%	32.11%	0.77%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%

present study, ethnic categories have been grouped because of small sample sizes. In Tables 1 and 2, which combine information across the 14 programs, the "Asian" heading is a relabeling of category 3, the "Hispanic" heading includes categories 4, 7, and 8, and the "Other and missing" heading includes category 1, as well as those who are missing ethnic information. In the tables for individual programs (A1-A28), further collapsing of ethnic categories was necessary. Information is provided for Whites, Blacks, and foreign students; all other categories are included under the "Other and missing" heading. In interpreting Tables 1-2 and A1-A28, it is important to know that for the earlier years of data (through approximately 1976), Northwestern sometimes omitted ethnic codes for Whites. It is not possible to distinguish these White students from students for whom ethnic codes were omitted for other reasons. This explains the higher percentage of missing data and the lower percentage of Whites in the earlier years.

The most striking aspect of the information in Tables 1-2 and A1-A28 is the increase in the percentage of foreign students in most programs. Overall, the percentage of entering Ph.D.-seekers who were foreign increased from 15% to 32%. The most dramatic changes were the increases in Computer Science, from 28% foreign students in 1975-1977 to 62% in 1984-1986, and in Physics, from 29% to 60%. (The large percentage change in Theatre, from 4% to 40%, is less noteworthy because of the small number of students.) Increases in the percentage of foreign students have been evident in other studies as well (e.g., National Research Council, 1986; Trent and Copeland, 1987).

The percent of Black enrollees dropped from 3.3 to 1.6; the percent of Hispanics dropped from 3 to 1. The percent of Asians was less than 1 in 1975-77, reached 4 in 1981-1983, and dropped to 3 in 1984-1986. (Note that these percentages of minority enrollment do not include foreign students.)

Combined across programs, the 2:1 ratio of men to women has remained quite steady, although the ratio of men to women varies considerably across programs. The most significant within-program changes over time were the increases in the proportion of women in Clinical Psychology and Counseling Psychology. There was a large decrease in the proportion of women in Theatre.

Tables 3 through 6 provide three types of information about Ph.D.-seeking students: (1) the percentage of students who attained candidacy by the end of the data collection in May, 1987 (2) the percentage of those attaining candidacy who also graduated and (3) the overall percentage of students who graduated. Admission to Ph.D. candidacy at Northwestern is contingent on completion of departmental requirements, including a comprehensive qualifying examination, and on the approval of the Graduate Faculty.

Northwestern's Graduate School has regulations concerning the amount of time permitted for achieving candidacy and graduation. These official timetables must be considered in interpreting our findings, although,

Table 3

Percents of 1972 - 1978 Entrants
Attaining Graduate School Milestones by May, 1987:
Results for the 14 Graduate Programs

Program	Sample Size	Candidacy (%)	Graduation, given Candidacy (%)	Graduation (%)
Counseling Psychology	126	73	82	60
Chemistry	193	83	96	80
English	100	46	72	33
History	80	61	67	41
Mathematics	62	48	73	36
Political Science	104	51	77	39
Chemical Engineering	80	59	96	56
Clinical Psychology	63	84	85	71
Economics	148	54	85	46
Philosophy	50	62	58	36
Physics	78	58	96	55
Sociology	91	67	82	55
Theatre	53	25	46	11
Computer Science	151	31	83	26
Total	1379	59	83	49

Table 4

Percents of 1975-1978 Entrants Who Attained Candidacy By May 1987: Results for Gender and Ethnic Groups[a]

	White	Black	Foreign	Other and Missing[b]	Total
Male	64%	50%	62%	22%	58%
	(380)	(6)	(111)	(78)	(575)
Female	64%	68%	65%	23%	57%
	(188)	(19)	(26)	(52)	(285)
Missing Gender	---	---	---	6%	6%
	(0)	(0)	(0)	(17)	(17)
Total	64%	64%	63%	20%	56%
	(568)	(25)	(137)	(147)	(877)

[a]Sample sizes are shown in parentheses.

[b]Hispanics, Asians, and Native Americans are included in this category.

Table 5

Percents of 1975-1978 Entrants Who Graduated by May, 1987, Given That They Achieved Candidacy: Results for Gender and Ethnic Groups[a]

	White	Black	Foreign	Other and Missing[b]	Total
Male	81%	67%	86%	82%	82%
	(242)	(3)	(69)	(17)	(331)
Female	72%	77%	100%	83%	76%
	(121)	(13)	(17)	(12)	(163)
Missing Gender	---	---	---	0%	0%
	(0)	(0)	(0)	(1)	(1)
Total	78%	75%	88%	80%	80%
	(363)	(16)	(86)	(30)	(495)

[a]Sample sizes are shown in parentheses.

[b]Hispanics, Asians, and Native Americans are included in this category.

Table 6

Percents of 1975-1978 Entrants Who Graduated by May, 1987: Results for Gender and Ethnic Groups[a]

	White	Black	Foreign	Other and Missing[b]	Total
Male	52% (380)	33% (6)	53% (111)	18% (78)	47% (575)
Female	46% (188)	53% (19)	65% (26)	19% (52)	44% (285)
Missing Gender	--- (0)	--- (0)	--- (0)	0% (17)	0% (17)
Total	50% (568)	48% (25)	55% (137)	16% (147)	45% (877)

[a]Sample sizes are shown in parentheses.

[b]Hispanics, Asians, and Native Americans are included in this category.

according to our results, they may not always have been followed. According to the 1985-1986 Northwestern catalog,

> a student is expected to be admitted to candidacy before the end of the third calendar year after initial registration in the Graduate School at Northwestern university; a student must be admitted to candidacy by the end of the twelfth quarter after initial registration... (Northwestern University, 1985, p. 34).

> All requirements for the doctoral degree must be met within five years of admission to candidacy, or within eight years of the last year of consecutive full-time residency, to be calculated from the beginning of that year, or within ten years of the initial registration in the Graduate School, whichever comes first...A student may petition for a [two-year] extension of the deadline... There is no extension beyond two years (p.33).

Table 3 gives candidacy and graduation information by graduate program for students who entered during the years 1972-1978. Collapsing across programs, the rates for candidacy; graduation, given candidacy; and graduation were 59%, 83%, and 49%, respectively. The highest candidacy and graduation rates were in Clinical Psychology (84% candidacy, 71% graduation) and Chemistry (83% candidacy, 80% graduation), while the lowest rates were in Theatre (25% candidacy, 11% graduation) and Computer Science (31% candidacy, 26% graduation). The highest rates for graduation, given that candidacy had been attained, were in Chemistry, Chemical Engineering, and Physics (96% in each case); the lowest were in Theatre (46%) and Philosophy (58%). Northwestern staff have informed us that in some programs, students who are in reality seeking only a master's degree may state that they are seeking a Ph.D. in order to make themselves eligible for certain types of financial aid. This may, in part, explain the low rates of attainment in Computer Science and Theatre.

Tables 4 through 6 give information for ethnic and gender groups, combined across the 14 graduate programs, for students who entered during the years 1975-1978. (The ethnic categories are defined as in Tables A1-A28.) Collapsing across groups, the rates of candidacy; graduation, given candidacy; and graduation were 56%, 80%, and 45%, respectively. These rates differ slightly from those reported above, which were based on students who entered from 1972-1978. As shown in Table 4, candidacy rates for Whites, Blacks, and foreign students were nearly identical (63% to 64%). Only for Black students was there a substantial difference in the candidacy rates for males (50%) and females (68%), but because of the small sample sizes (n = 6 for Black males), this finding should not be given too much weight. As indicated in Table 5, the rate of graduation, given candidacy, was higher for foreign students (88%) than for Whites (78%) and Blacks (75%). Among Whites, the rate was higher for men (81%) than for women (72%). Among foreign

students and Blacks, however, the rate was higher for women. (Again, note the small sample size for Blacks.) All 17 female foreign students who achieved candidacy also graduated. The pattern of ethnic and gender differences was the same for the graduation rates, shown in Table 6. The rate for foreign students (55%) exceeded rates for Whites (50%) and Blacks (48%). Among Whites, the rate was higher for men, whereas among Black and foreign students, the rates were higher for women.

There are several possible reasons for the higher rates of graduation for foreign students. Foreign students are likely to have been selected to study in the United States because of their academic excellence. Also, as Girves and Wemmerus (in press, p. 10) pointed out, "the fact that foreign students must be enrolled full-time and must demonstrate sufficient financial support to carry out their degree programs may be more incentive for them to complete their degrees. Domestic students, on the other hand, do not necessarily have these incentives, and may have other options outside of graduate school."

The very low candidacy and graduation rates for students who were missing gender or ethnic information (see Tables 4 and 6) is somewhat mysterious. (Although the "Other and missing" ethnic category includes Asians, Hispanics, and Native Americans, about 80% of the students in this category were, in fact, missing ethnic information.) It could be that when students drop out, there is less opportunity for university personnel to fill in missing information on gender or ethnicity, thus creating an association between low attainment and the absence of these data. Because of the coding practices mentioned earlier, it is likely that a large proportion of students who are missing ethnic codes are, in fact, White. If all the students who were missing ethnic data were White, the rates of candidacy and graduation for Whites would be roughly 6 to 8 percentage points lower. The rates for Whites given in Tables 4 and 6 can be viewed as upper bounds on the actual rates.

Patterns of Attainment of Graduate School Milestones

The rates of attainment given in Tables 4-6 can provide only limited information about patterns of candidacy and graduation. A more detailed picture, based on all entering students, rather than 1972-1978 entrants only, can be achieved through survival analysis, a method often used in biostatistical applications (Kalbfleisch and Prentice, 1980). In survival analysis, we are interested in the survival function, which is the probability that an event will take more than x units of time to occur. In this study, the events of interest are the graduate school milestones, graduation and candidacy, and the units of time are graduate school years. The survival function, S(x), is defined as follows:

$$S(x) = P(X > x) = 1 - F(x),$$

where X is the time elapsed until the milestone is reached and F(x) is the cumulative distribution function of X. A related function is the hazard

function, h(x), which is the instantaneous risk of the occurrence of an event at X = x, given that the event has not occurred before time x. The hazard function is defined as

$$h(x) = f(x) / S(x),$$

where f(x) is the probability density function of X. The hazard and survival functions are equivalent ways of summarizing the distribution of survival times, since

$$h(x) = - d/dx [\ln (S(x))].$$

If X is assumed to have the exponential distribution with parameter θ, h(x) is constant and equal to θ.

It is important to note that the statistical terms, "survival," "hazard," and "risk," are used here in a way that differs from everyday parlance. In our report, survival refers to the probability of remaining in graduate school without achieving the event of interest; for example, the probability that the degree is not received by a particular time. Similarly, we speak of the "hazards" or "risks" of attaining candidacy or completing a degree. (Some may find this usage to be counterintuitive in the present context; others may find it appropriate!)

Standard methods exist for both nonparametric and parametric estimation of survival functions (e.g., see Kalbfleisch and Prentice, 1980). Difficulties in estimation can occur when sample sizes are small, however, particularly when it is of interest to estimate separate curves for subpopulations. Bayesian methods can yield more stable estimates by incorporating prior distributions for model parameters. Whereas previous Bayesian efforts have focused on the estimation of a single survival curve, Braun (1985) developed an empirical Bayes (EB) approach for estimating a family of survival functions. Details of the model and the estimation procedures are provided in Appendix B. A general description of EB methods is given in Braun (in press).

Three types of survival analyses were conducted. The first two types pertained to the achievement of candidacy and graduation, respectively. The third type of analysis involved examination of the attainment of the Ph.D. degree, given that candidacy had been reached. For each of the three types of analysis, graphs of the EB estimates of the hazard and survival functions are provided.

To facilitate interpretation of the survival analysis graphs, the 14 selected graduate programs have been grouped as follows: Group I consists of the two programs that are part of the Technological Institute, Chemical Engineering and Computer Science, and the three most technical of the Arts and Sciences programs, Chemistry, Mathematics, and Physics. Group II consists of the three behavioral science programs, Counseling Psychology, Clinical Psychology, and Sociology, as well as the Theatre program, and Group III includes the remaining Arts and Sciences programs, English, History,

Political Science, Economics, and Philosophy. (Additional analyses of candidacy and graduation were conducted in which only White Americans were included. The EB analysis produced results similar to those obtained for the total group of students; the classical analysis produced more unstable results because of the smaller sample sizes.)

A phenomenon that had to be considered in analyzing these data is censoring: the removal of individuals from the risk set (the group of individuals who are available to experience the event of interest) for reasons other than the occurrence of the event. In this study, some individuals were censored because the data collection effort ended during their graduate careers. In the survival model applied here, censoring is accommodated through adjustment of the risk set. This means that if the termination of the data collection effort occurs at time x of a student's graduate career, that student will no longer be considered "exposed" or "at risk" for candidacy or graduation after time x. Note that, for purposes of our analyses, students who left graduate school without a Ph.D. are still considered to be part of the risk set. Roughly speaking, our analyses focused on the probabilities of achieving milestones in year x for those who entered school x years earlier. If it had been possible to obtain accurate information about student drop-out, the students who left school without attaining milestones could have been deleted from the risk set. This type of analysis, however, would have had a different interpretation. It would have involved estimation of the probabilities of attaining milestones by year x <u>for those students still in school x years after entry</u>. In an analysis of this kind, the attainment of milestones would have appeared more likely.

The results of the survival analyses for the 14 selected graduate programs are given in Figures 1-18. The initial sample sizes for the 14 graduate programs ranged from 76 to 414 for the unconditional analyses of candidacy and graduation and from 25 (for Theatre) to 281 for the analysis of graduation, given candidacy. (In survival analyses, the size of the risk set decreases as more people attain the event. Therefore, estimates of hazard and survival functions for later time periods are based on fewer cases and are less precise than those for earlier time periods.) Note that, for each function, the vertical and horizontal scales of the graphs are the same <u>within</u> each of the three types of analysis, but they differ somewhat across analysis types.

Figures 1-6 give the results of the candidacy analyses. For each of the three groups of programs, the graph of the estimated hazard functions appears first, followed by the graph of the estimated survival function. The hazard function at time x can be interpreted as the instantaneous "risk" that the event (candidacy or graduation) occurs at time x, given that it has not occurred prior to time x. The survival function at time x is the probability that the event has not occurred by time x. If the hazard function takes on a high value at time x, the survival function will show a correspondingly large drop at time x.

FIGURE 1

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR CANDIDACY

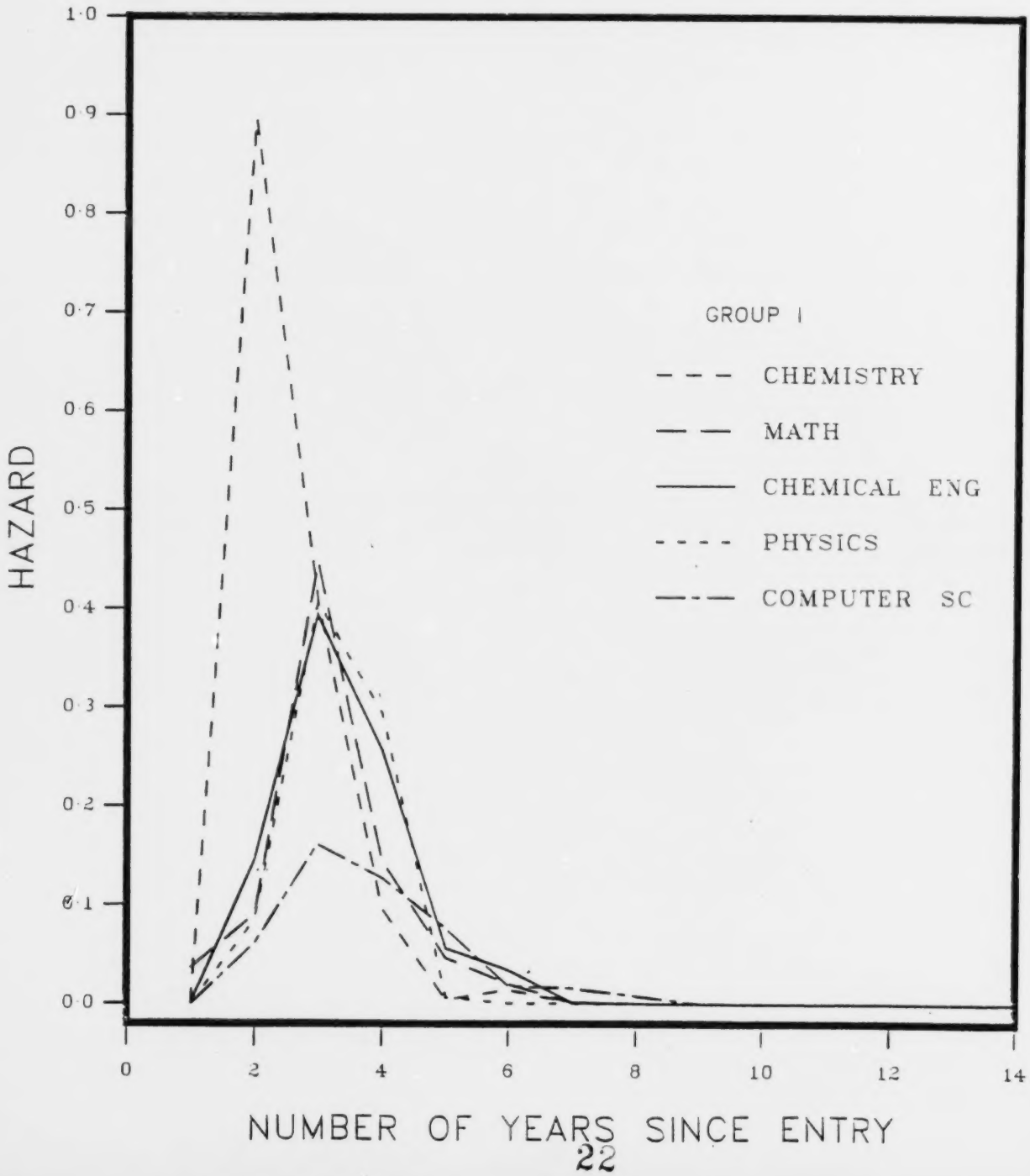


FIGURE 2

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR CANDIDACY

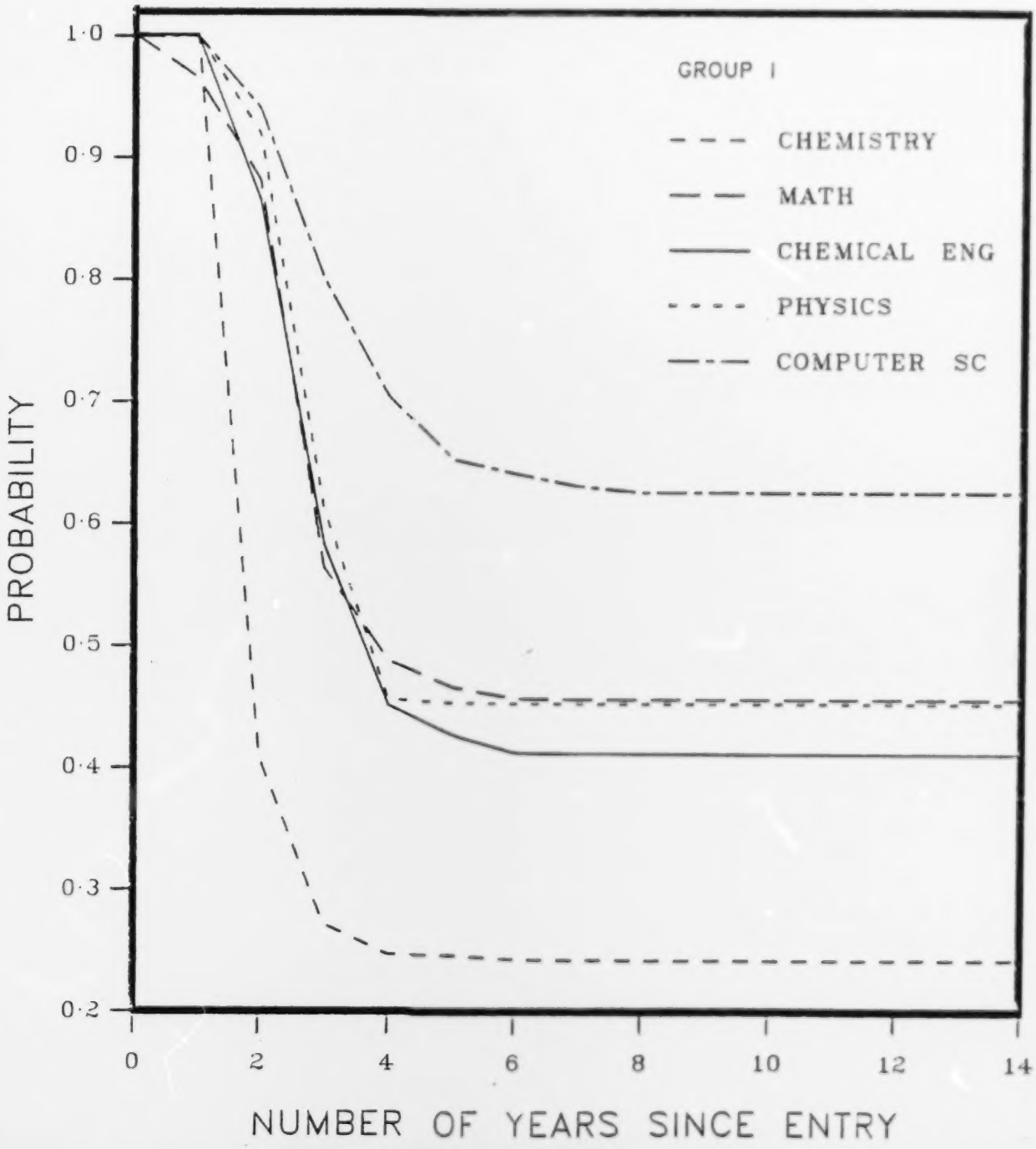


FIGURE 3

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR CANDIDACY

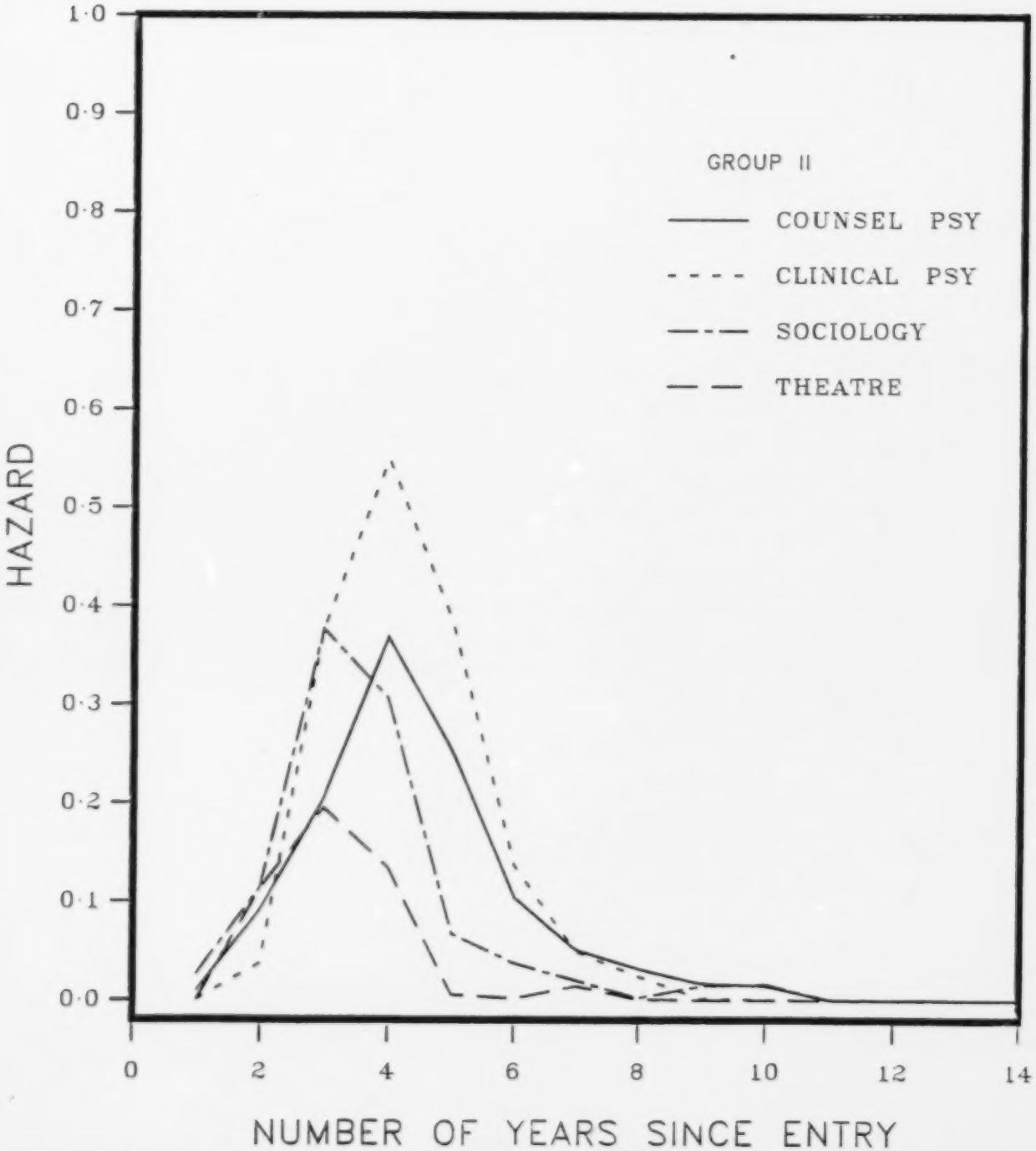


FIGURE 4

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR CANDIDACY

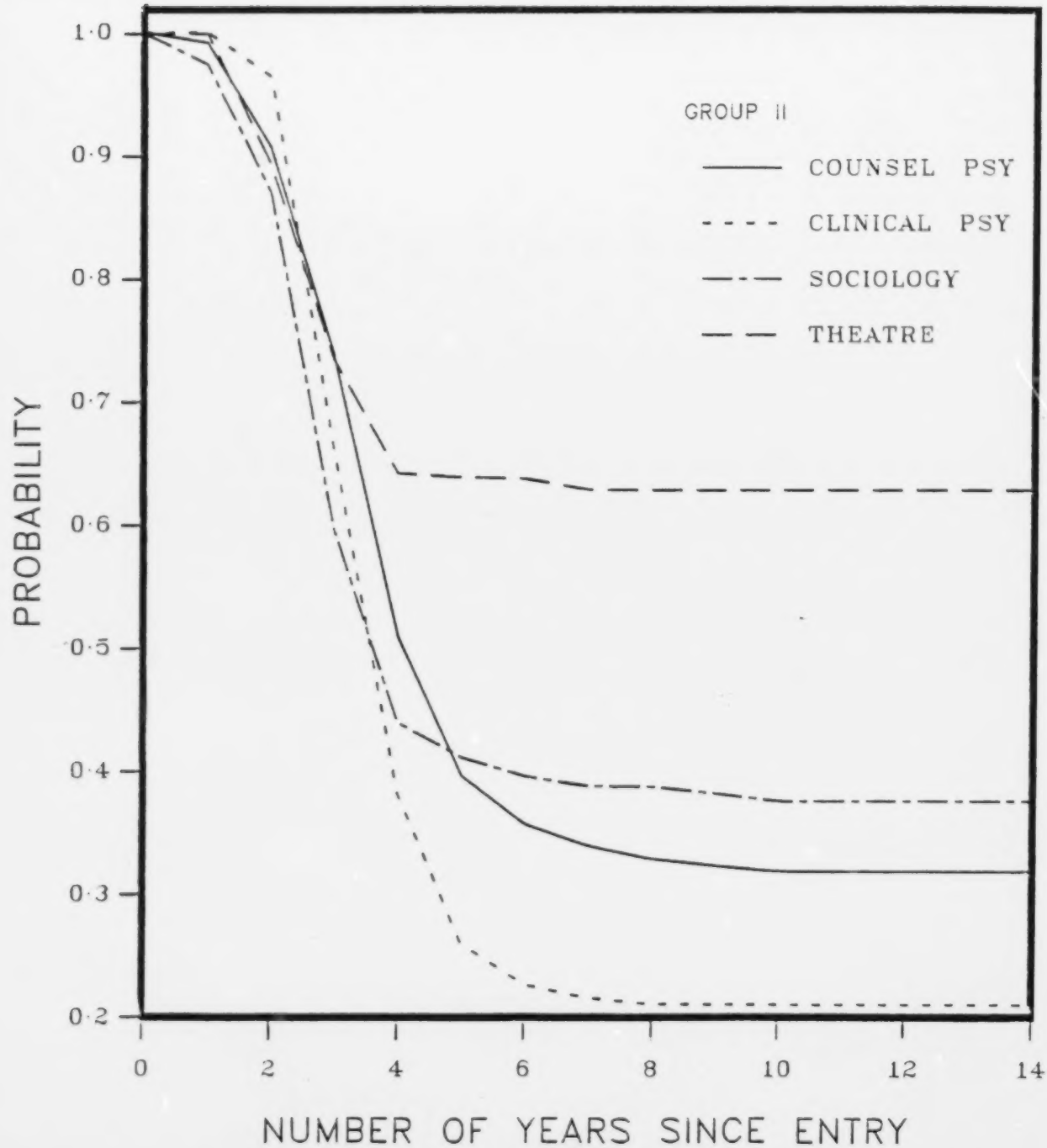


FIGURE 5

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR CANDIDACY

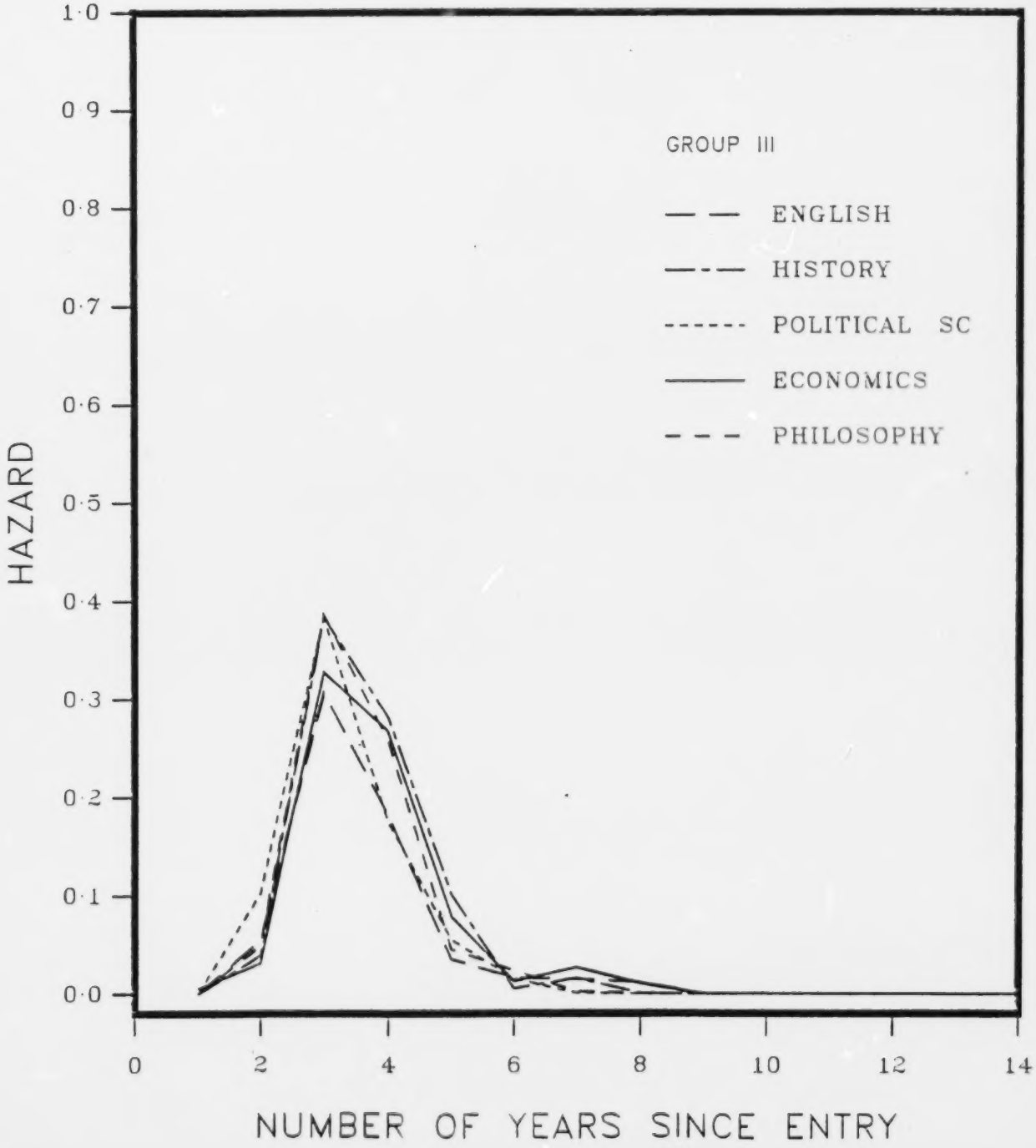
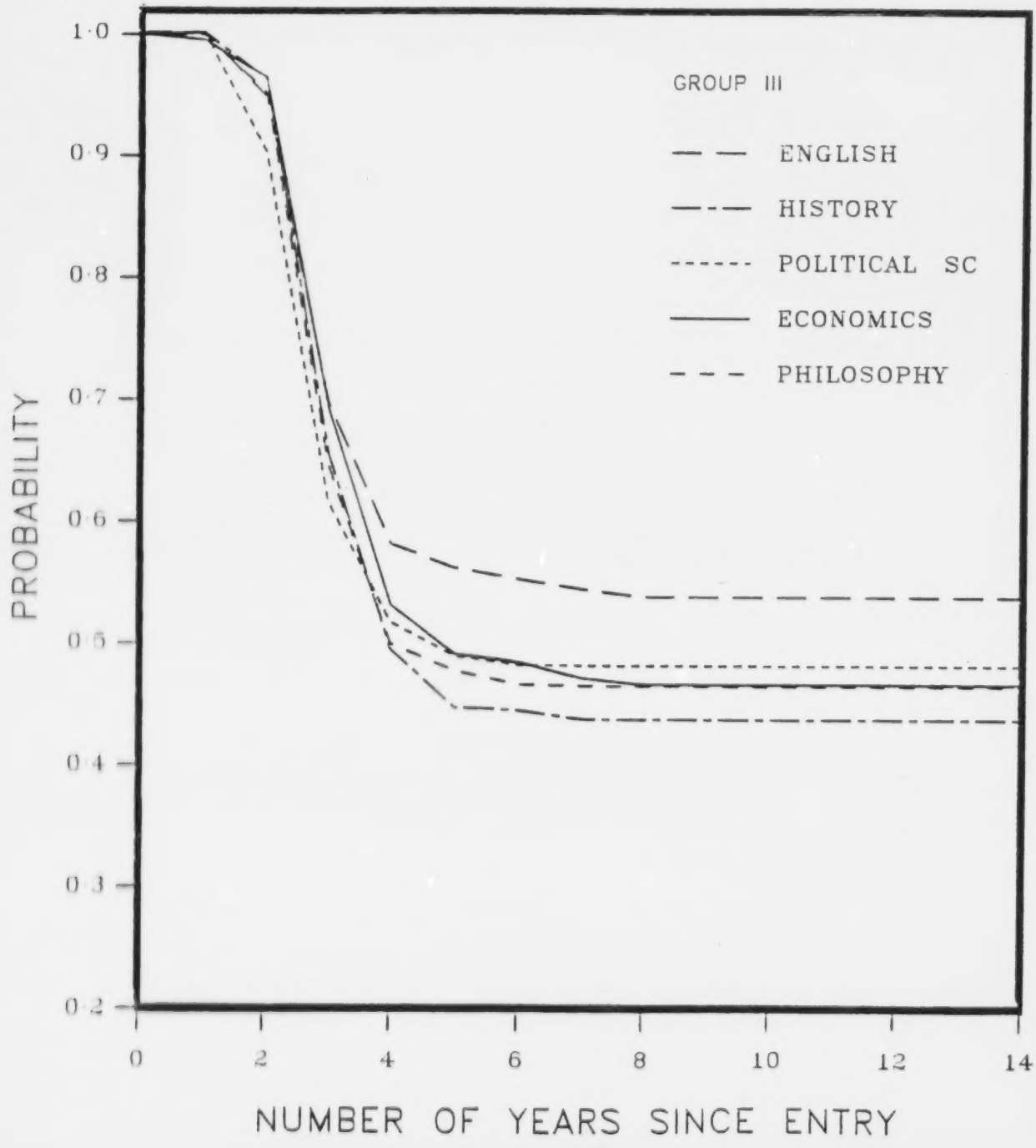


FIGURE 6

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR CANDIDACY



One interesting aspect of the candidacy analyses in Figures 1-6 is that results were similar among the programs in Group III, but not among the programs in Group I and II. In Group III (Figure 5), all five hazard functions rose to a sharp peak at year 3, and then declined, indicating that the third year of graduate school was the most likely time for the occurrence of candidacy in these programs. Correspondingly, the survival functions (Figure 6) showed a steep drop until year 4, and then started to level out. For four of the programs, the probability that candidacy is not achieved dropped to below .50 by year 5 and remained about the same through year 14. (English fell to about .55.) Apparently, if candidacy is not achieved by year 4, it is unlikely to be achieved. The hazard functions for Groups I and II (Figures 1 and 3) showed peaks in years 2-4, but there was substantial variation among programs in the shapes of the hazard functions. The survival functions (Figures 2 and 4) leveled out by about the fifth year, but the values they eventually reached varied widely, from about .65 for Theatre and Computer Science to slightly more than .20 for Clinical Psychology and Chemistry.

In the analyses of graduation, which are displayed in Figures 7-12, there was again greater similarity among the Group III programs than among Groups I and II. The hazards for Group III (Figure 11) peaked at year 5 except for History, which peaked at years 6 and 10. The survival functions (Figure 12) leveled out between years 10 and 12, reaching values between about .70 for Philosophy and .55 for Economics. As in the candidacy analyses, Groups I and II displayed considerably more variation. Chemistry and Chemical Engineering showed significant peaks in the hazard functions at year 5 (Figure 7) as did Clinical Psychology at year 6 (Figure 9), corresponding to steep drops in the survival functions (Figures 8 and 10). For Group I, the survival functions (Figure 8) leveled off by year 8, at values ranging from about .70 for Computer Science to about .30 for Chemistry. The Group II survival functions (Figure 10) leveled off by year 12 at values ranging from about .80 for Theatre to about .35 for Clinical Psychology.

Figures 13-18 show the results of the survival analysis for graduation, given that candidacy has occurred. In these figures, the x-axis represents years since the attainment of candidacy, rather than years since entry to graduate school. In these analyses, the Group I programs were the closest together and also showed the steepest drops. As of the sixth year after candidacy, the values of the survival function (Figure 14) ranged from about .25 for Math (representing a probability of .75 of completing a degree by this point for those who achieved candidacy) to slightly below .10 for Chemistry and Chemical Engineering. For Groups II and III (Figures 16 and 18), the programs with the highest estimated survival probabilities were Theatre, History and Philosophy (all about .50); the program with the lowest value was Clinical Psychology (.10).

It is hoped that the analyses of the type displayed in Figures 1-18 can be useful to graduate school deans in estimating the number of graduates an entering class is likely to yield and in determining whether administrative

FIGURE 7

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR GRADUATION

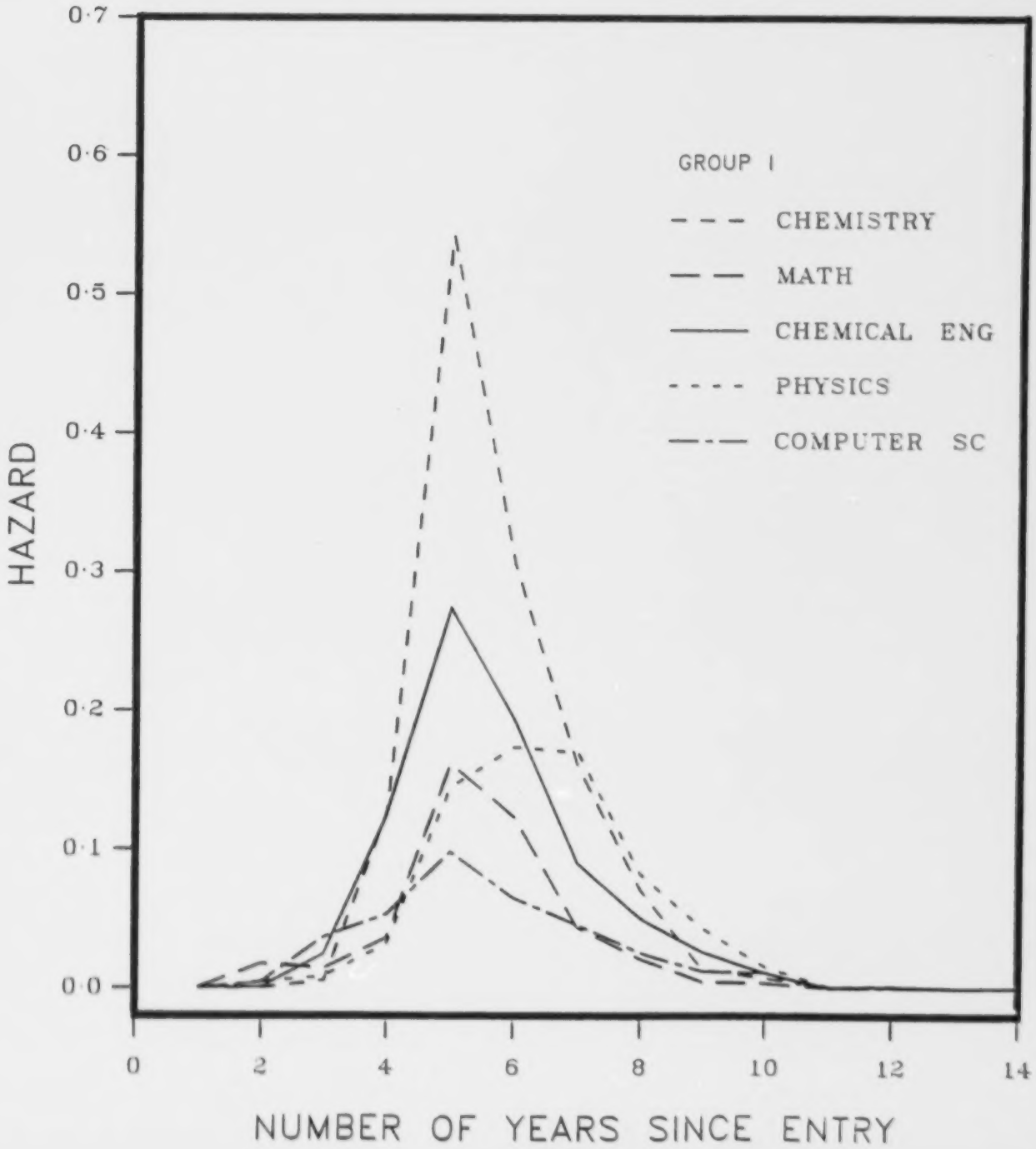


FIGURE 8

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR GRADUATION

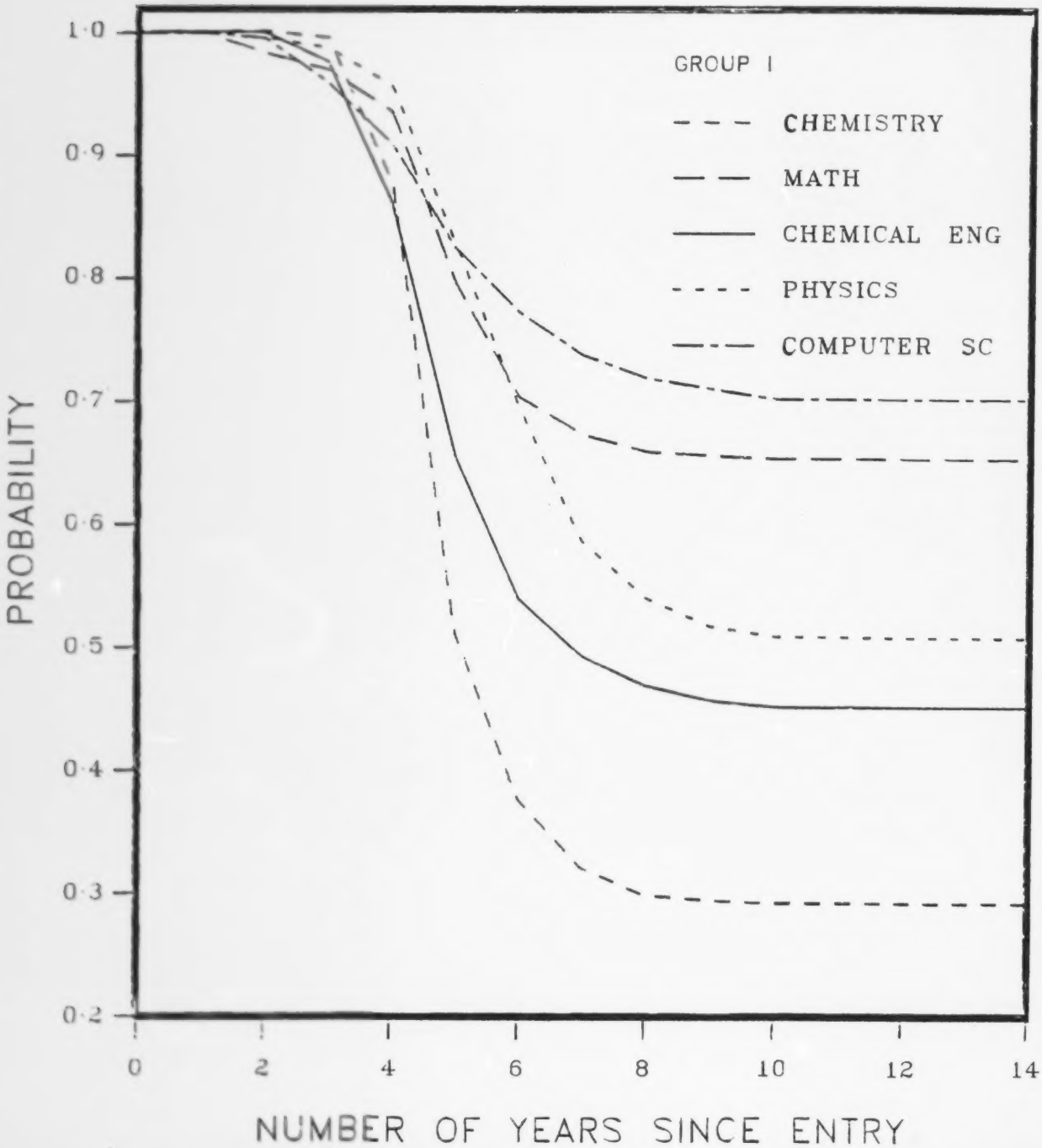


FIGURE 9

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR GRADUATION

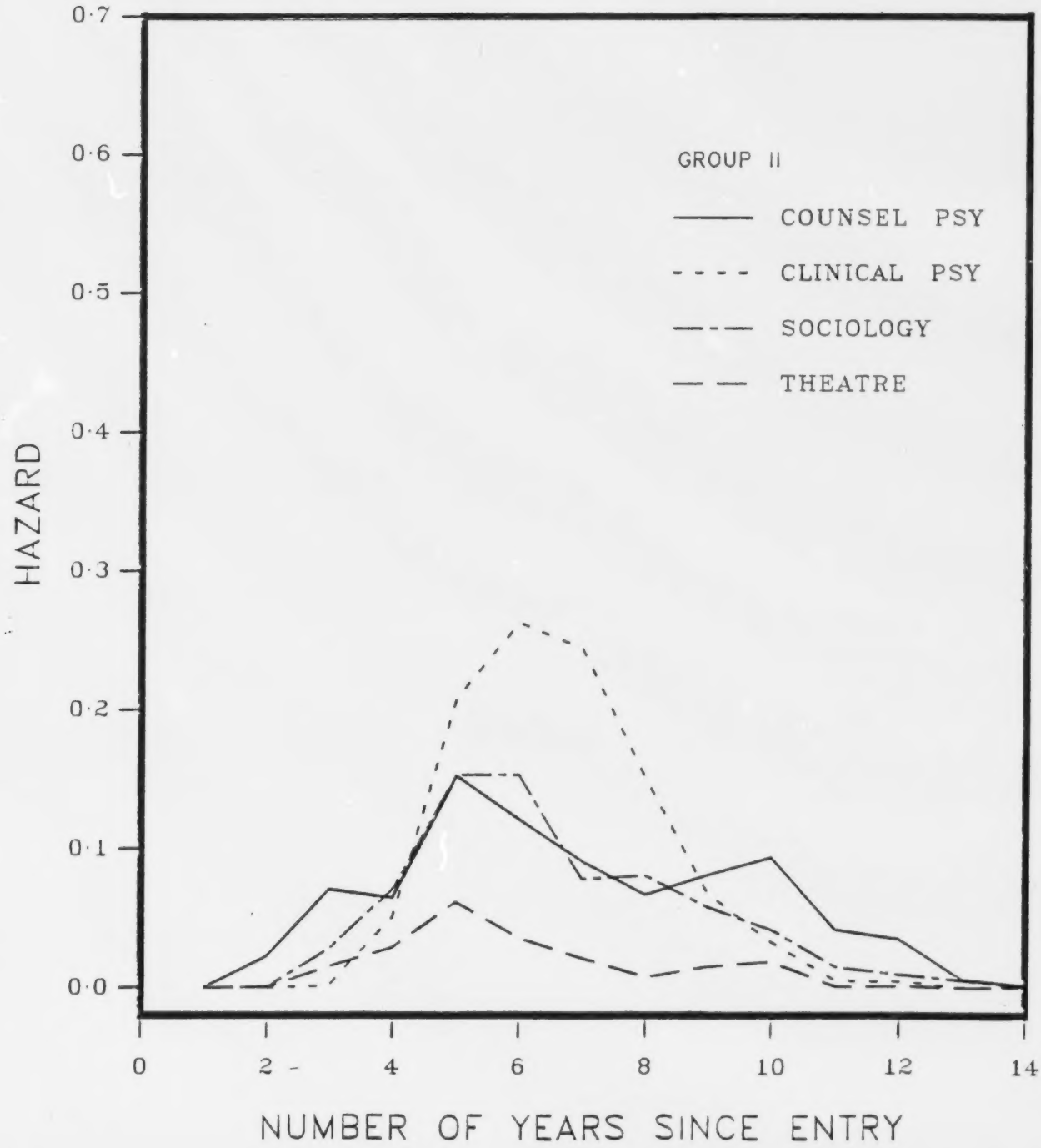


FIGURE 10

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR GRADUATION

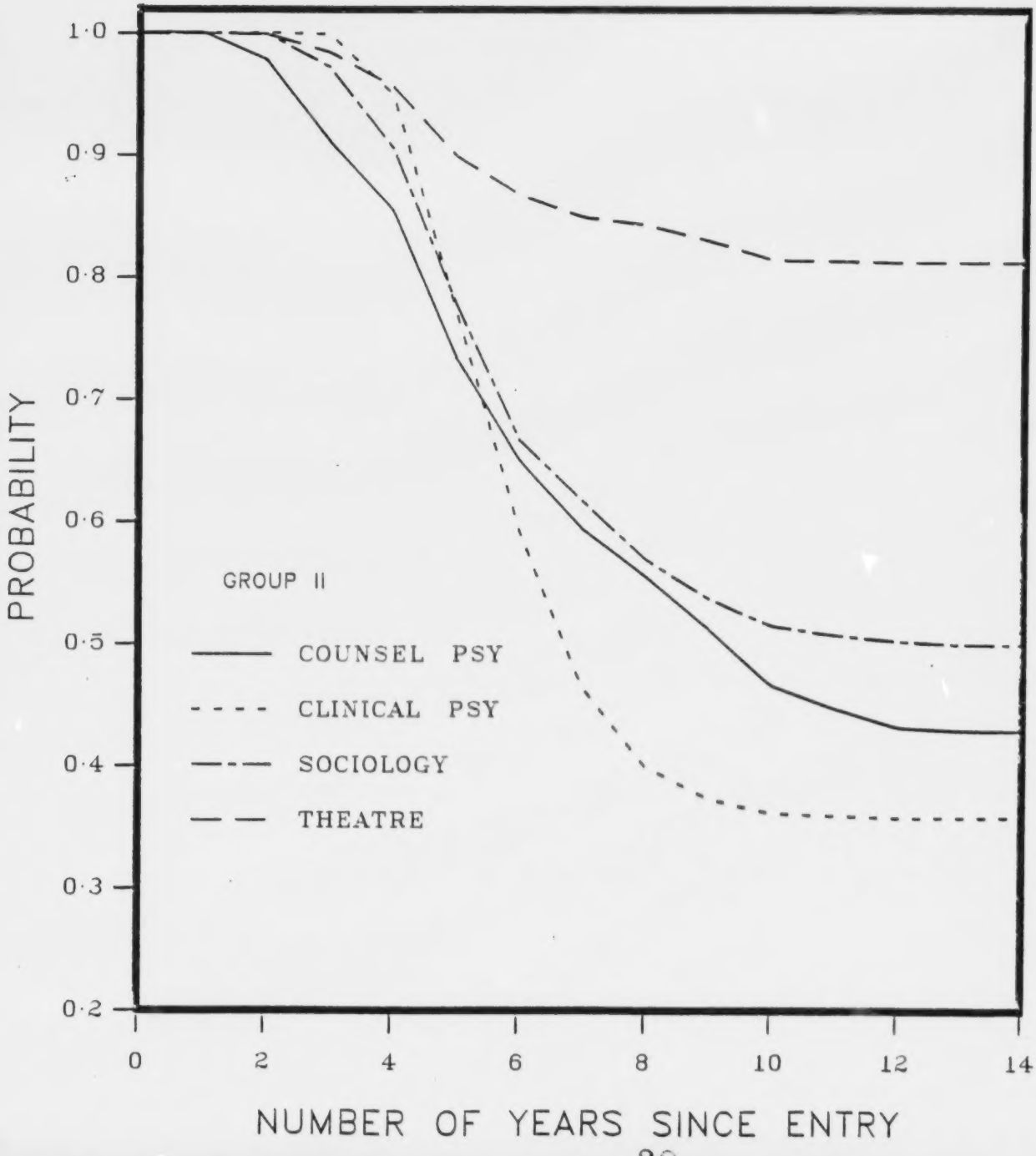


FIGURE 11

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR GRADUATION

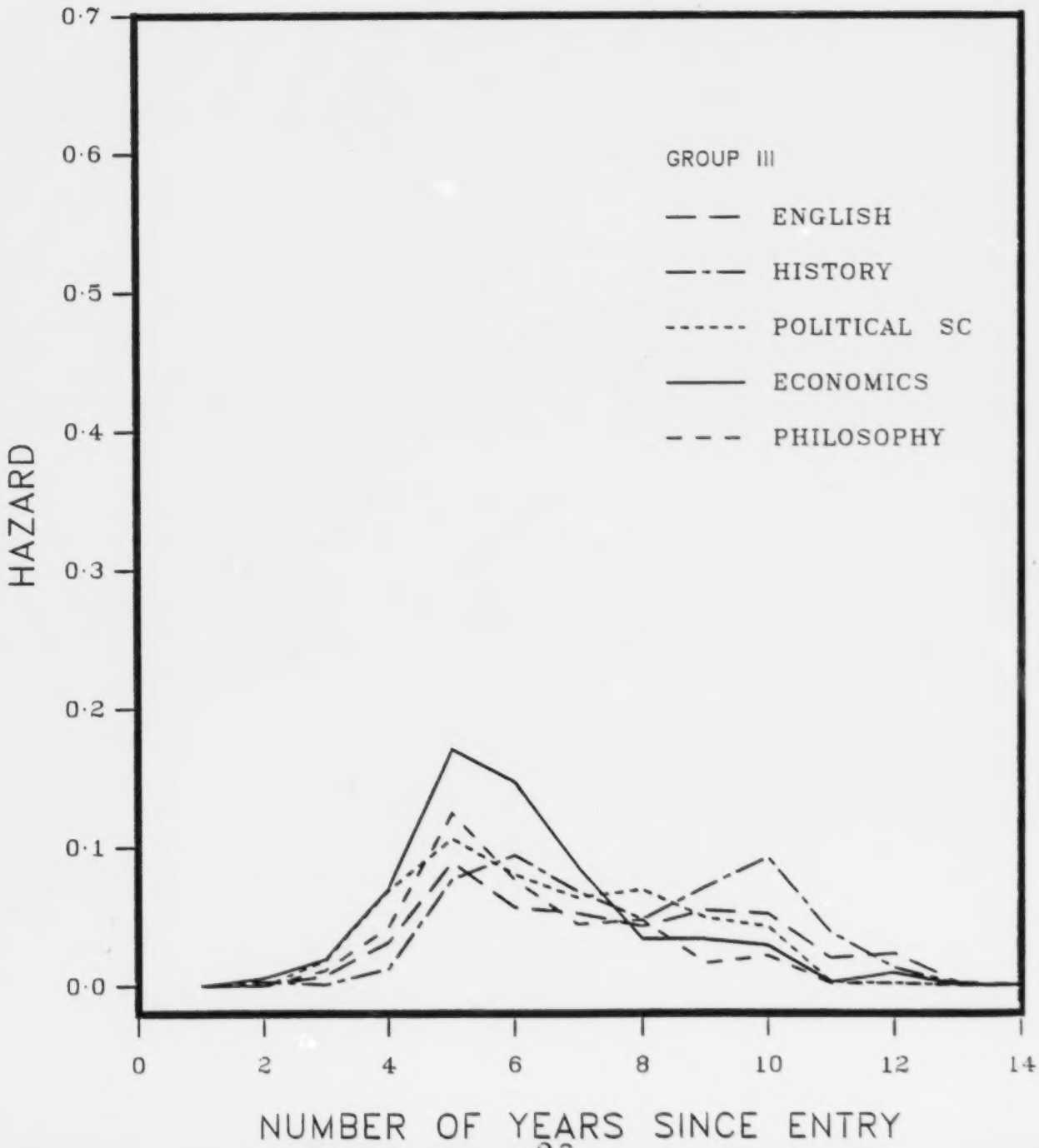


FIGURE 12

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR GRADUATION

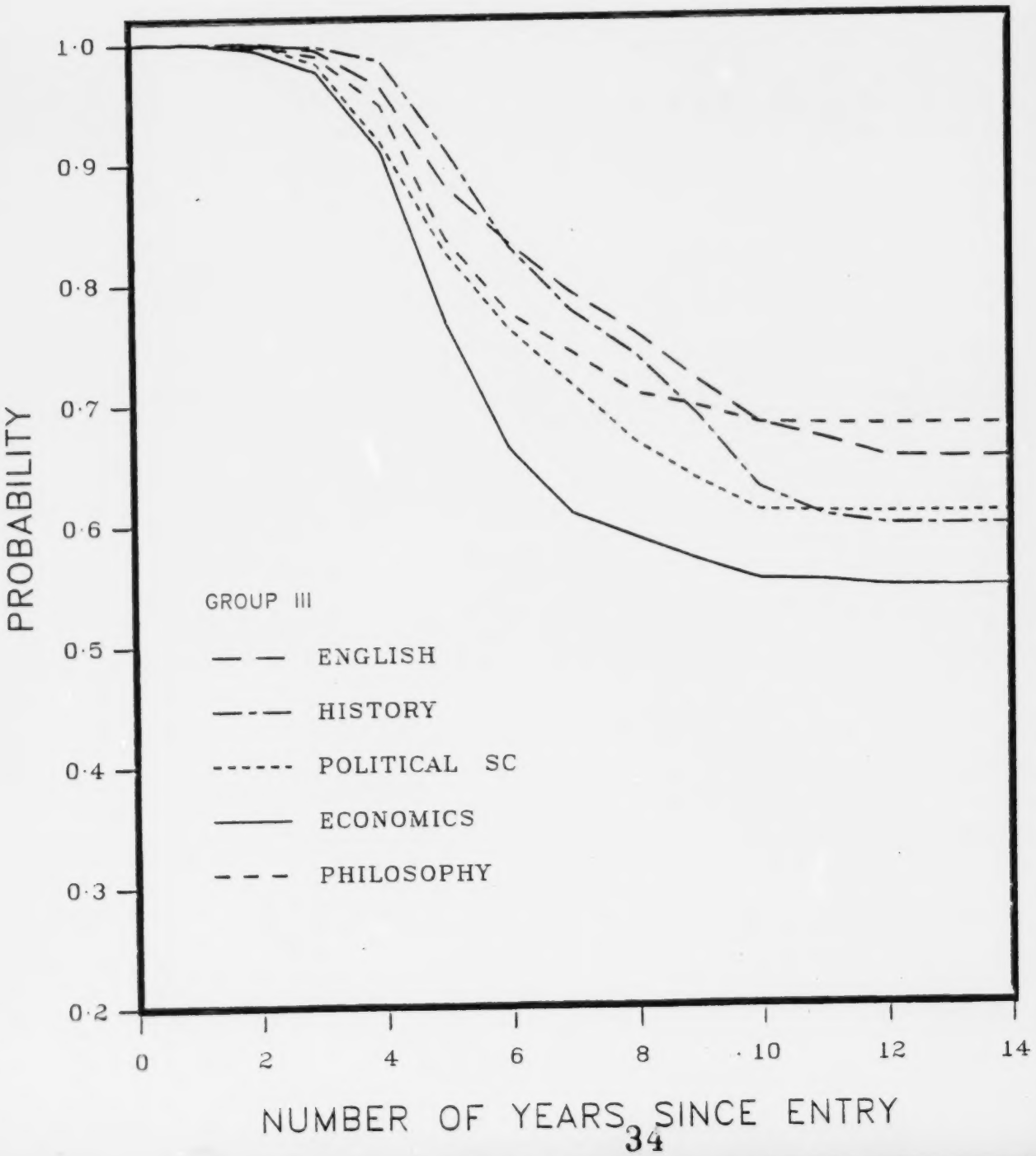


FIGURE 13

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR GRADUATION

PH.D. CANDIDATES

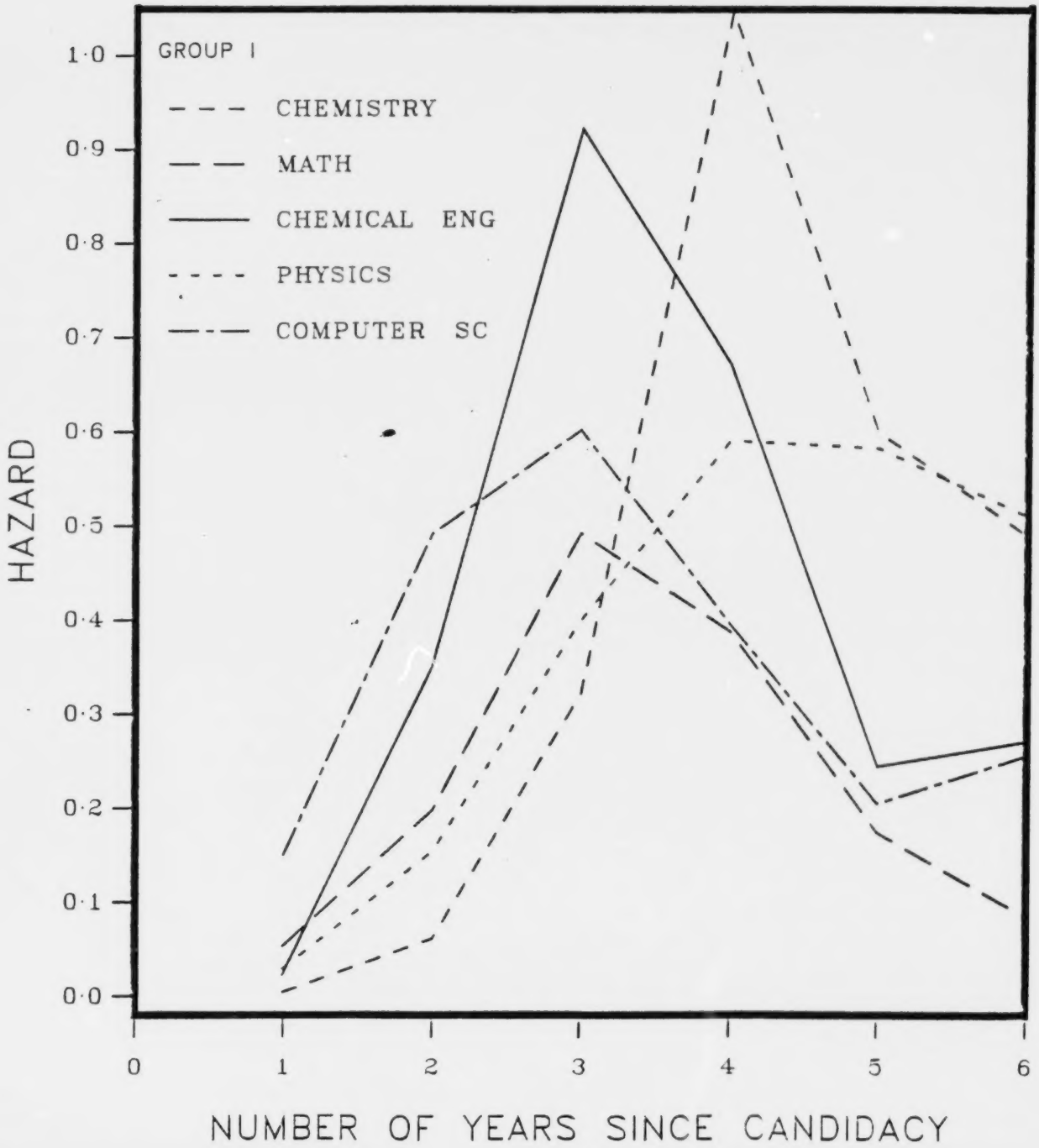


FIGURE 14

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR GRADUATION

PH.D. CANDIDATES

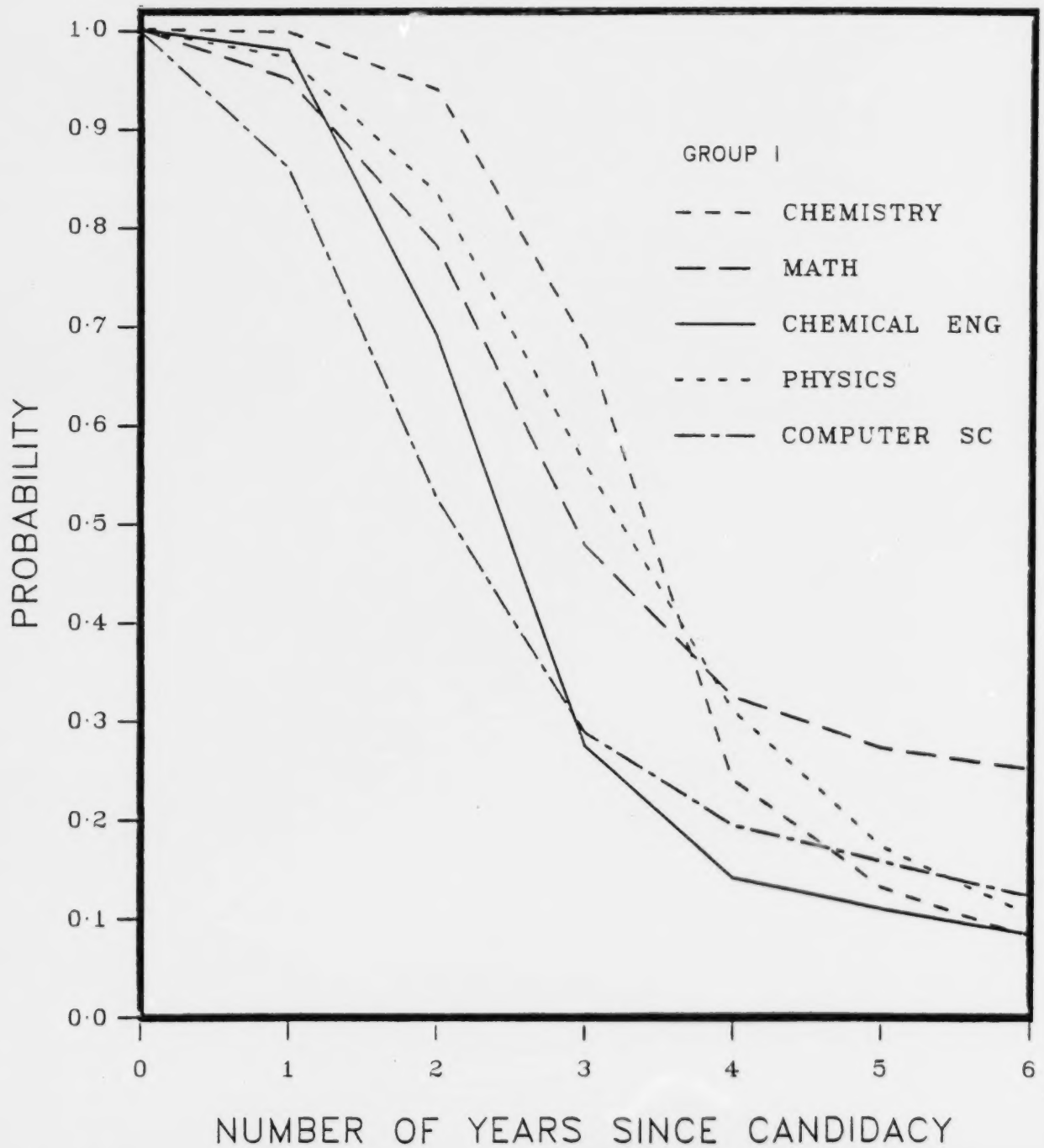


FIGURE 15

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR GRADUATION

PH.D. CANDIDATES

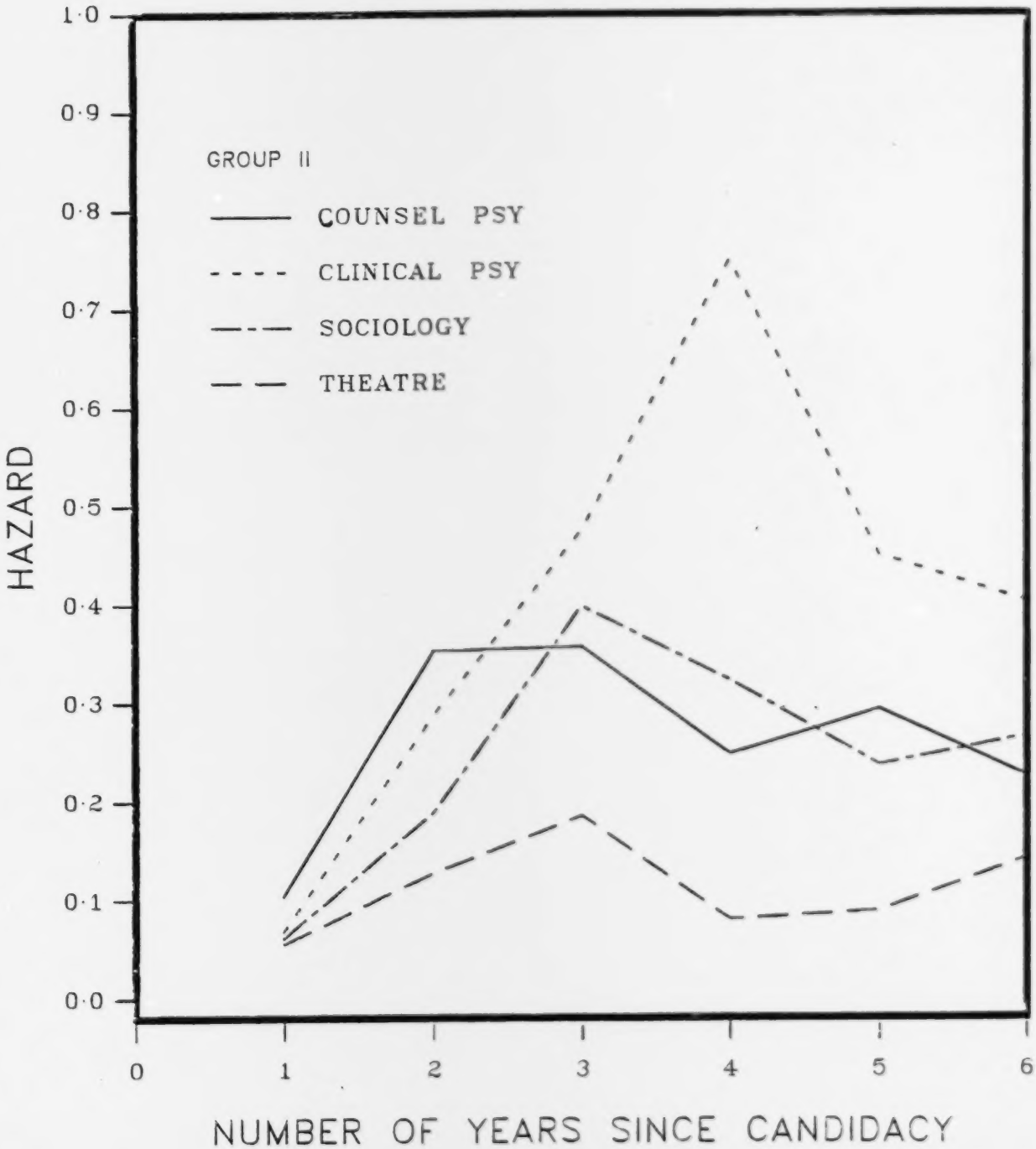


FIGURE 16

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR GRADUATION

PH.D. CANDIDATES

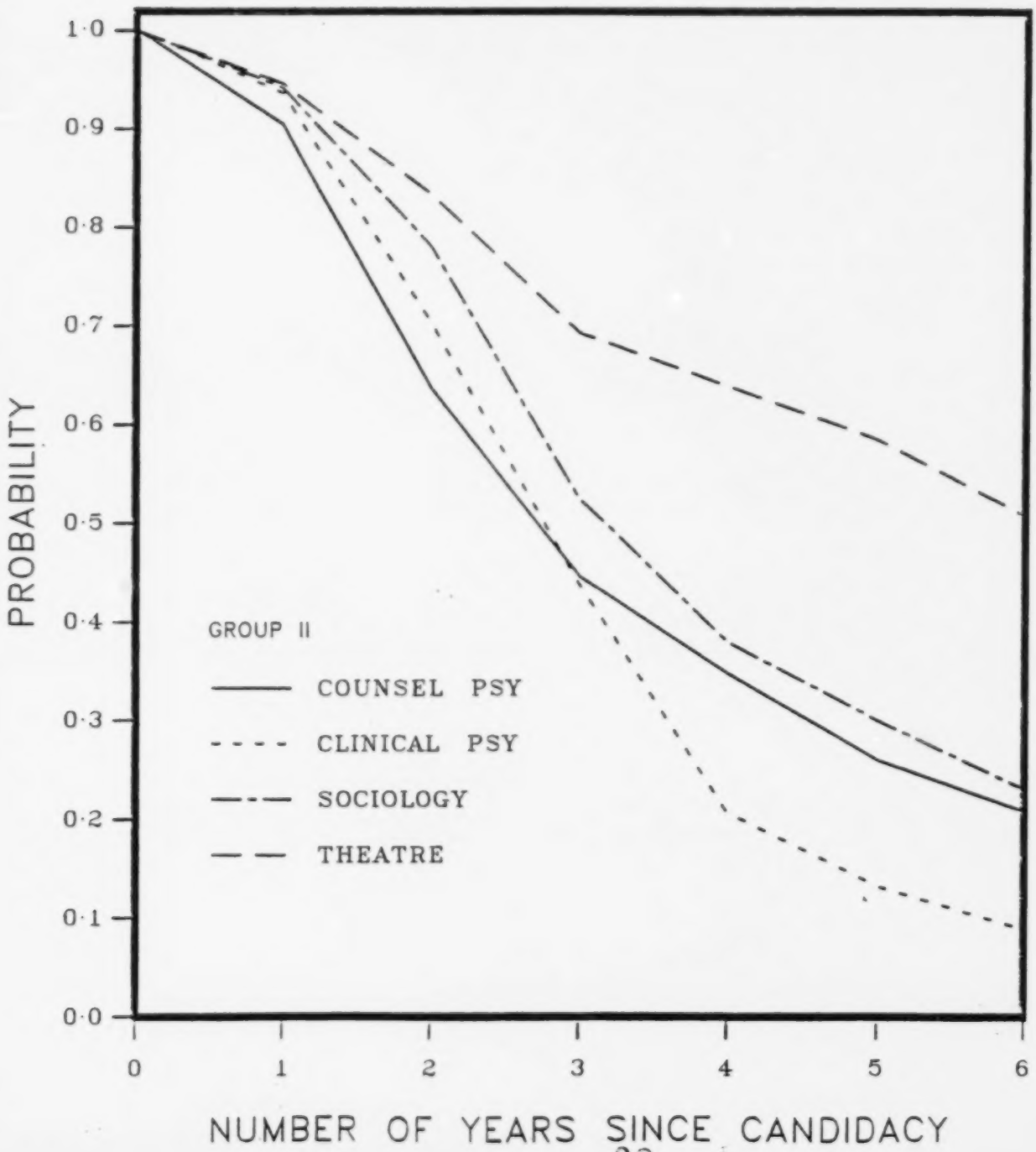


FIGURE 17

BAYES ESTIMATES OF HAZARD FUNCTIONS FOR GRADUATION

PH.D. CANDIDATES

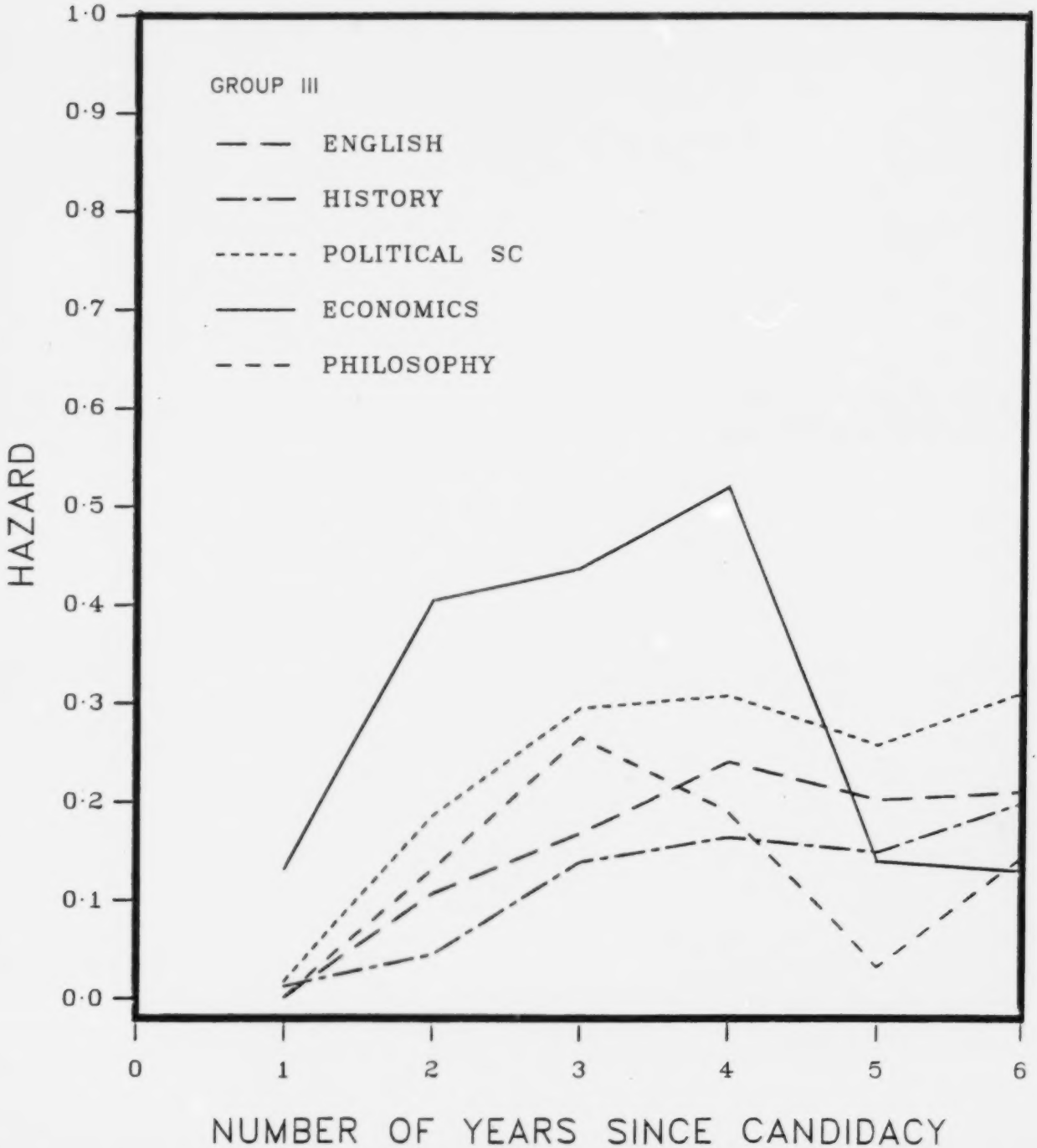
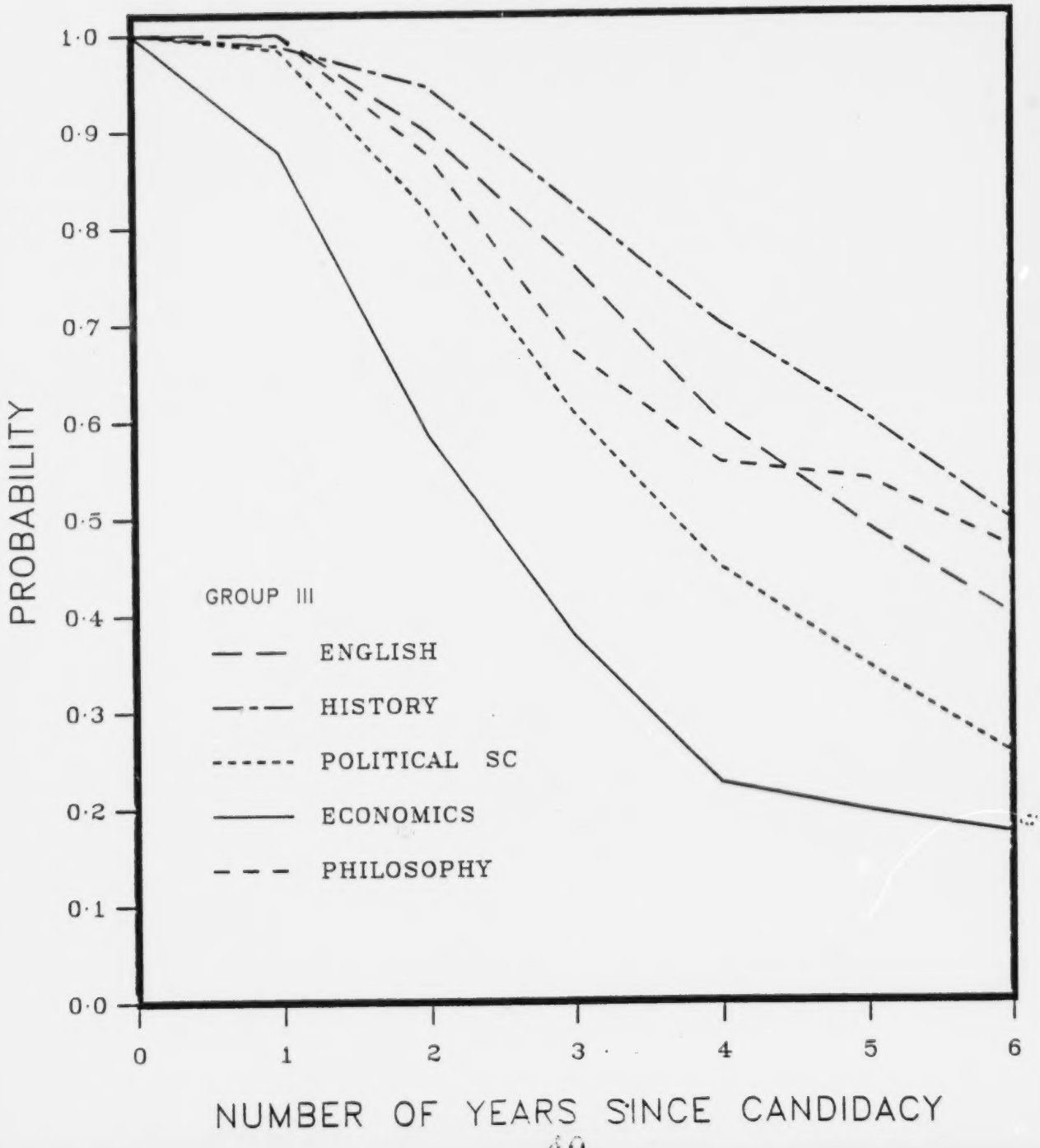


FIGURE 18

BAYES ESTIMATES OF SURVIVAL FUNCTIONS FOR GRADUATION

PH.D. CANDIDATES



changes are needed to hasten progress in some graduate programs. Survival analysis allows examination of candidacy and graduation rates at multiple time points and thus gives a more detailed picture of milestone attainment than simple rates of candidacy or graduation. For example, Figure 12 allows us to state that, of an entering class of 10 students in Political Science or Economics at Northwestern, one student would be expected to receive the doctorate by four years after entry. Analyses of the kind illustrated in Figures 13-18, which show the rate at which Ph.D. candidates complete their degrees, should be particularly useful to policymakers in targeting programs for administrative review.

Relation of Candidacy and Graduation to Measures of Academic Potential

Our original intention was to use logistic regression analysis (see Hanushek and Jackson, 1977) to model the relation between milestone attainment and such explanatory variables as undergraduate grade-point average (UGPA), GRE verbal score (GREV), and GRE quantitative score (GREQ). A possible EB strategy for logistic regression, which takes advantage of existing EB methods for the normal case, is outlined in Appendix C.

However, preliminary examination of the data revealed that GRE scores and UGPA were almost entirely unrelated to the achievement of candidacy and graduation. The candidacy and graduation variables were defined as follows: Individuals received a code of one if they attained the milestone by August, 1986 and a code of zero otherwise. (That is, both dropouts and those who remained in school without attaining the milestone received a code of zero.) Only students who entered between 1972 and 1978 were included in the analysis. GRE scores were available for 76% of these students overall; percents ranged from 48 to 94 across graduate programs. UGPA was available for 84%, with percents ranging from 58 to 95. Means and standard deviations of GRE scores and UGPA for the 14 graduate programs are given in Table 7, along with the percent of 1972-1978 entrants for which predictor information was available. (Scores were not available for the GRE analytical measure, which was first administered in its present form in 1981.)

Graphical displays of the correlations of the candidacy and graduation variables with GRE scores and UGPA are given in Figures 19-20. Figure 19 shows the point-biserial correlations between the candidacy indicator variable and GREV, GREQ, and UGPA. (A point-biserial correlation is a Pearson correlation between a dichotomous variable and a continuous variable; see, e.g., McNemar, 1962.) The left-most column lists the intervals for values of the correlation coefficients. The next column shows, for each interval, two-letter codes for the graduate programs for which the correlation between GREV and candidacy fell in that interval. (Graduate programs are listed alphabetically within intervals.) The next two columns give the analogous information for the correlations of candidacy with GREQ and UGPA, respectively. Figure 20 shows the corresponding correlations for the graduation indicator variable. The sample sizes on which these correlations are based ranged from 43 to 172. For some students, information was available for some preadmissions measures, but not others. Typically, correlations involving GREV and GREQ were based on identical or nearly

Table 7

Descriptive Statistics for Predictor Variables
(1972-1978 Entrants)

Graduate Program	N	Graduate Record Examination Verbal Mean	Verbal SD	Quantitative Mean	Quantitative SD	%[a]	Undergraduate GPA Mean	GPA SD	GPA %[a]
Counseling Psychology	126	599	93	543	106	48	3.21	.41	94
Chemistry	193	584	86	714	69	84	3.49	.34	89
English	100	712	80	583	108	94	3.56	.35	92
History	80	643	107	556	127	86	3.58	.38	91
Math	62	593	141	731	83	74	3.62	.34	84
Political Science	104	584	117	588	115	79	3.49	.37	78
Chemical Engineering	80	460	107	695	92	65	3.39	.41	59
Clinical Psychology	63	650	81	636	93	83	3.53	.39	95
Economics	148	623	114	702	76	81	3.50	.40	74
Philosophy	50	697	69	666	89	94	3.66	.31	90
Physics	78	553	143	722	64	76	3.32	.41	58
Sociology	91	612	113	584	123	78	3.50	.34	86
Theatre	53	604	104	547	109	81	3.46	.33	91
Computer Science	151	523	140	706	79	58	3.47	.39	75

[a]Percent of 1972-1978 entrants for which predictor information was available.

Figure 19

Point-Biserial Correlations of Measures of Academic Attainment with Candidacy Indicator Variable[ab]

Interval[c]	GRE-Verbal	GRE-Quantitative	Undergraduate GPA
[-.20, -.15)		CP,PC	PH, SO
[-.15, -.10)		HI	HI, TH
[-.10, -.05)	CH, EC, SO		CH, PS
[-.05, 0)	CS, HI	CE	CP
[0, .05)	PC	EC, SO	CS, EC, PC
[.05, .10)	PH, PS	PS, TH	
[.10, .15)	CL, EN	CH, EN, MA	MA
[.15, .20)	CE, CP	CL, PH	CL
[.20, .25)		CS	CE, EN
[.25, .30)	MA, TH		
Median	.06	.05	-.01

Key:

CE - Chemical Engineering	EC - Economics	PH - Philosophy
CH - Chemistry	EN - English	PS - Political Science
CL - Clinical Psychology	HI - History	SO - Sociology
CP - Counseling Psychology	MA - Mathematics	TH - Theatre
CS - Computer Science	PC - Physics	

[a] The indicator variable equals one if candidacy was achieved by August, 1986 and zero otherwise. Only students who entered between 1972 and 1978 were included in the analysis.

[b] Sample sizes range from 43 to 172.

[c] Programs are listed alphabetically within intervals.

Figure 20

Point-Biserial Correlations of Measures of Academic Attainment with Graduation Indicator Variable[ab]

Interval[c]	GRE-Verbal	GRE-Quantitative	Undergraduate GPA
[-.30, -.25)		CP	
[-.25, -.20)			SO
[-.20, -.15)	HI		HI
[-.15, -.10)	EC	HI	TH
[-.10, -.05)	PS	PC	
[-.05, 0)	CH, CS, PC		CH, EC, PC, PH
[0, .05)	CP, PH	CE, EC, MA, PS	CP, CS, PS
[.05, .10)	EN	EN	
[.10, .15)	CE, CL, SO	PH, SO, TH	CL, EN, MA
[.15, .20)	MA	CH	
[.20, .25)		CS	CE
[.25, .30)		CL	
[.30, .35)			
[.35, .40)	TH		
Median	.03	.06	-.00

Key:

CE - Chemical Engineering	EC - Economics	PH - Philosophy
CH - Chemistry	EN - English	PS - Political Science
CL - Clinical Psychology	HI - History	SO - Sociology
CP - Counseling Psychology	MA - Mathematics	TH - Theatre
CS - Computer Science	PC - Physics	

[a] The indicator variable equals one if graduation was achieved by August, 1986 and zero otherwise. Only students who entered between 1972 and 1978 were included in the analysis.

[b] Sample sizes range from 43 to 172.

[c] Programs are listed alphabetically within intervals.

identical groups of students; analyses involving UGPA were based on a slightly different, but overlapping, group.

For the candidacy variable, correlations ranged from -.19 to .29; for graduation, they ranged from -.23 to .37. The size of the correlations for GREV and GREQ seemed to be unrelated to the degree of quantitative emphasis in the graduate programs. For each of the three measures of academic potential, the ordering of the correlations in the graduation analysis roughly paralleled that obtained in the candidacy analysis. There were four programs -- Chemical Engineering, Clinical Psychology, English, and Math -- in which at least five out of the six correlations displayed in Figures 19-20 were positive and one program -- History -- in which all six correlations were negative. The most striking aspect of these results, however, is that, as shown in the last row of Figures 19 and 20, the medians for all six types of correlations were close to zero. Correlations of GREV, GREQ, and UGPA with reciprocal time to candidacy and reciprocal time to degree also tended to be very low, as did correlations of GRE advanced test scores with candidacy and graduation. It was hypothesized that GRE scores and UGPA might be more successful as predictors of graduation, given that candidacy had been achieved. Therefore, correlations of GREV, GREQ, and UGPA with graduation were computed for only those students who had achieved candidacy; these correlations, too, had medians close to zero. Finally, analyses were repeated for White Americans only, again producing similar results.

To facilitate further exploration of the interrelationships between measures of academic potential and the attainment of graduate school milestones, a listing of several key variables was obtained for those who had matriculated in any of the following eight programs during the years 1972 to 1978: Chemical Engineering, Chemistry, Mathematics, Counseling Psychology, Clinical Psychology, English, History, and Political Science. (The earlier phases of this study included these eight programs only.) The following variables were listed: sex, ethnicity, UGPA, GREV, GREQ, a weighted sum of UGPA, GREV, and GREQ, the candidacy and graduation indicator variables, and the number of years to completion of the Ph.D., where applicable. These data were examined in detail and were tabulated in various ways. For example, stem-and-leaf diagrams of UGPA, GREV, GREQ, and the composite variable were created for those who had and had not achieved candidacy and graduation. Males, females, Blacks, and Whites were examined separately. These painstaking analyses were intended to reveal any patterns that had might have gone undetected in more conventional analyses. However, no such patterns were found.

In typical validity studies of the GRE, researchers examine the correlation of GRE scores with grade-point average for the first year of graduate school (see Burton & Turner, 1983). First-year GPA was not available in the Northwestern data base. For students who completed doctorates, we examined the correlations of final GPA with GRE scores and UGPA. Unfortunately, the number of graduates within each program for whom predictor information was available was very small. Therefore, the sample sizes for these correlations averaged about 27. As in previous analyses, the correlations for the three predictors are based on somewhat different subsets of students. The median correlations for the three predictors were .05 for

GREV, .25 for GREQ, and .35 for UGPA. The correlations of UGPA with final graduate school GPA were always positive, unlike the correlations involving GRE scores, which were negative in 4 out of 14 programs for both GREV and GREQ.

These results show that, in a population of Ph.D.-seeking matriculants in Northwestern's graduate school, conventional measures of verbal and quantitative skills cannot discriminate between students who do and do not achieve candidacy and graduation. This does not, of course, imply that the GRE and UGPA are not useful in admissions: The population of graduate school matriculants has already been selected on the basis of GRE scores, UGPA, and other factors and those with the least potential for achieving candidacy or graduation are likely to have been weeded out. Therefore, the low correlations are not unexpected (see Dawes, 1975; Rubin, 1980). (In a summary of previously conducted studies of the relation between GRE scores and Ph.D. attainment, Willingham, 1974, reported median correlations of .18 for GREV and .26 for GREQ. These results are not directly comparable to the present findings because the 47 correlations on which each median was based came from different institutions and corresponded to different administrative units.) In the case of the correlations between preadmissions measures and final GPA at Northwestern, selection is even more severe, since only those who completed graduate school are included in the analysis. The within-program means for final GPA ranged from 3.50 to 3.90, with standard deviations typically less than .25.

Within the select population of graduate school matriculants, it is likely that personality factors such as perseverance, as well as the availability of financial and social support, play a crucial role in determining whether graduate school milestones are attained. In a study that included a student survey, Girves and Wemmerus (in press) found that involvement in the graduation program (e.g., participation in research projects, seminars, meetings, and social activities), student relationships with faculty, and financial support had a direct or indirect effect on progress toward the doctoral degree. There is some evidence that, at the undergraduate level, admissions test results and preadmissions grades also have little association with persistence toward the degree: Willingham (1985) obtained the biserial correlations between a composite of high school rank and SAT and persistence to the senior year of college. These correlations were found to be very low; in six of the nine colleges studied, they did not reach statistical significance.

These findings suggest that further research on candidacy and graduation rates should focus on noncognitive factors. It may be that improvements in candidacy and graduation rates can best be achieved by designing admissions procedures that place more weight on personality attributes like determination or persistence and by improving support systems for students already in school.

Summary

Several types of analyses were conducted, based on about 2700 Ph.D.-seeking students who matriculated in 14 programs at Northwestern University's

Graduate School during a 15-year period. Descriptive analyses of students in these programs who entered between 1975 and 1986 showed that the percentage of foreign students increased from 15 in 1975-1977 to 32 in 1984-1986. The percentages of Blacks and Hispanics dropped from about 3 to 1 during this time, while the percentage of Asians increased from 1 to 3. Combined across programs, the ratio of men to women remained relatively steady at about 2:1.

Combined across all 14 graduate programs, the rates of candidacy and graduation for students with at least eight years of opportunity to achieve these milestones were 59% and 49%, respectively. The rate of graduation, given that candidacy had occurred was 83%. There were substantial variations in these rates across graduate programs and, to a lesser degree, across demographic groups. The highest candidacy and graduation rates were in Clinical Psychology and Chemistry; the lowest were in Theatre and Computer Science. The rate of candidacy was the same for Whites, Blacks, and foreign students, but both types of graduation rates were higher for foreign students than for Blacks and Whites. Among Whites, graduation rates were higher for men; among foreign students, they were higher for women. The superiority of candidacy and graduation rates for Black women over those for Black men must be interpreted with caution because of small sample sizes. Interpretation of the ethnic group results is complicated by the absence of ethnic codes for about 17% of the students in the analysis.

Survival analyses of candidacy and graduation showed that the Group III programs -- English, History, Political Science, Economic, and Philosophy -- produced very similar patterns. For candidacy, the survival functions showed a steep drop to roughly .50 (corresponding to a candidacy rate of 1 - .50 = .50) in year 4 and then started to level out. For graduation, most Group III programs showed sharp drops in their survival curves at about year 5. The survival functions leveled out between years 10 and 12 to values between .55 and .70 (corresponding to graduation rates between .45 and .30). Survival functions for Groups I and II showed a great deal of variation. For example, the values at which the survival functions for graduation leveled out ranged between about .30 for Chemistry and .80 for Theatre.

Analyses of the relation between measures of academic potential, such as GREV, GREQ, and UGPA, with candidacy and graduation showed little relationship between preadmission measures and milestone attainment. Most of the within-program correlations ranged between -.25 and .25; the medians of these correlations across the 14 programs were close to zero for each of the six pairs of variables. Evidently, within this select group of students, these conventional measures of academic skills cannot discriminate between those who do and do not achieve candidacy and graduation.

The current study does not, of course, provide any information as to whether the obtained results may be generalized beyond Northwestern University. A multi-institution study is now underway that will involve investigation of some of the phenomena examined here, with a particular focus on the graduate school careers of minority students.

References

Braun, H. I. (in press). Empirical Bayes methods: A tool for exploratory analysis. In R. D. Bock and L. Burstein (Eds.), Proceedings of a conference on multilevel analysis of educational data (Princeton, NJ, 1987). New York: Academic Press.

Braun, H. I. (1985). Bayesian analysis of families of survival curves. Unpublished paper, Educational Testing Service.

Braun, H. I., Jones, D. H., Rubin, D. B., & Thayer, D. T. (1983). Empirical Bayes estimation of coefficients in the general linear model from data of deficient rank. Psychometrika, 48, 171-181.

Brown, S. V. (1987). Minorities in the graduate education pipeline. [A research report of the Minority Graduate Education project, jointly sponsored by the GREB and ETS]. Princeton: Educational Testing Service.

Burton, N. W., & Turner, N. J. (1983). Effectiveness of the Graduate Record Examinations for predicting first-year grades: 1981-82 summary report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.

Dawes, R. M. (1975). Graduate admissions variables and future success. Science, 187, 721-723.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, B, 39, 1-38.

Girves, J. E., & Wemmerus, V. (in press). Developing models of graduate student progress. Journal of Higher Education.

Hanushek, E. A., & Jackson, J. E. (1977). Statistical methods for social scientists. New York: Academic Press.

Hartnett, R. T. (1987). Has there been a graduate student "brain drain" in the arts and sciences? Journal of Higher Education, 58, 562-585.

Kalbfleisch, J. D., & Prentice, R. L. (1980). The statistical analysis of failure time data. New York: Wiley.

Korn, E. L., & Whittemore, A. S., (1979). Methods for analyzing panel studies of acute health effects of air pollution. Biometrics, 35, 795-802.

Laird, N. M. & Olivier, D. (1981). Covariance analyses of censored survival data using log-linear analysis techniques. Journal of the American Statistical Association, 76, 231-240.

McNemar, Q. (1962). Psychological Statistics (3rd ed.). New York: Wiley.

References (Continued)

Northwestern University (1985). The Graduate School, Vol. VIII, No. 2. Evanston, Illinois: Northwestern University.

National Research Council (1986). Summary report 1984: Doctorate recipients from United States Universities. Washington, DC: National Academic Press.

Rubin, D. B. (1980). Using empirical Bayes techniques in the law school validity studies. Journal of the American Statistical Association, 75, 801-816.

Szatrowski, T. H. (1976). Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and non-linear structure. (Technical Report No. 107.) Stanford, CA: Stanford University Statistics Department.

Trent, W. T., & Copeland, E. J. (1987). Effectiveness of state financial aid in the production of Black doctoral recipients. Atlanta: Southern Education Foundation.

Willingham, W. W. (1974). Predicting success in graduate education. Science, 183, 273-278.

Willingham, W. W. (1985). Success in college: The role of personal qualities and academic ability. New York: College Entrance Examination Board.

Wong, G. Y. & Mason W. M. (1985) The hierarchical logistic regression model for multilevel analysis. Journal of the American Statistical Association, 80, 513-524.

Appendix A

Tables A1 - A28

Ethnic and Gender Composition
of 1975 - 1986 Entrants

Table A1

Ethnic Composition for 1975-1986 Entrants[a]:
Chemical Engineering

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	16	0	14	10	40
	ROW%	40.00%	0.00%	35.00%	25.00%	100.00%
	COL%	26.23%	0.00%	23.73%	50.00%	28.37%
REGYR 1978-80	N	11	0	21	0	32
	ROW%	34.38%	0.00%	65.63%	0.00%	100.00%
	COL%	18.03%	0.00%	35.59%	0.00%	22.70%
REGYR 1981-83	N	14	0	13	5	32
	ROW%	43.75%	0.00%	40.63%	15.63%	100.00%
	COL%	22.95%	0.00%	22.03%	25.00%	22.70%
REGYR 1984-86	N	20	1	11	5	37
	ROW%	54.05%	2.70%	29.73%	13.51%	100.00%
	COL%	32.79%	100.00%	18.64%	25.00%	26.24%
TOTAL	N	61	1	59	20	141
	ROW%	43.26%	0.71%	41.84%	14.18%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A2

Proportions of Men and Women for 1975-1986 Entrants:
Chemical Engineering

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	32	4	4	40
	ROW%	80.00%	10.00%	10.00%	100.00%
	COL%	26.02%	28.57%	100.00%	28.37%
REGYR 1978-80	N	32	0	0	32
	ROW%	100.00%	0.00%	0.00%	100.00%
	COL%	26.02%	0.00%	0.00%	22.70%
REGYR 1981-83	N	28	4	0	32
	ROW%	87.50%	12.50%	0.00%	100.00%
	COL%	22.76%	28.57%	0.00%	22.70%
REGYR 1984-86	N	31	6	0	37
	ROW%	83.78%	16.22%	0.00%	100.00%
	COL%	25.20%	42.86%	0.00%	26.24%
TOTAL	N	123	14	4	141
	ROW%	87.23%	9.93%	2.84%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%

Table A3

Ethnic Composition for 1975-1986 Entrants[a]:
Computer Science

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	35	0	22	21	78
	ROW%	44.87%	0.00%	28.21%	26.92%	100.00%
	COL%	33.65%	0.00%	18.18%	70.00%	30.35%
REGYR 1978-80	N	22	0	19	1	42
	ROW%	52.38%	0.00%	45.24%	2.38%	100.00%
	COL%	21.15%	0.00%	15.70%	3.33%	16.34%
REGYR 1981-83	N	12	0	16	5	33
	ROW%	36.36%	0.00%	48.48%	15.15%	100.00%
	COL%	11.54%	0.00%	13.22%	16.67%	12.84%
REGYR 1984-86	N	35	2	64	3	104
	ROW%	33.65%	1.92%	61.54%	2.88%	100.00%
	COL%	33.65%	100.00%	52.89%	10.00%	40.47%
TOTAL	N	104	2	121	30	257
	ROW%	40.47%	0.78%	47.08%	11.67%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A4

Proportions of Men and Women for 1975-1986 Entrants:
Computer Science

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	63	9	6	78
	ROW%	80.77%	11.54%	7.69%	100.00%
	COL%	29.30%	25.00%	100.00%	30.35%
REGYR 1978-80	N	41	1	0	42
	ROW%	97.62%	2.38%	0.00%	100.00%
	COL%	19.07%	2.78%	0.00%	16.34%
REGYR 1981-83	N	29	4	0	33
	ROW%	87.88%	12.12%	0.00%	100.00%
	COL%	13.49%	11.11%	0.00%	12.84%
REGYR 1984-86	N	82	22	0	104
	ROW%	78.85%	21.15%	0.00%	100.00%
	COL%	38.14%	61.11%	0.00%	40.47%
TOTAL	N	215	36	6	257
	ROW%	83.66%	14.01%	2.33%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%

Table A5

Ethnic Composition for 1975-1986 Entrants[a]:
Chemistry

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	70	2	6	11	89
	ROW%	78.65%	2.25%	6.74%	12.36%	100.00%
	COL%	25.93%	66.67%	15.00%	50.00%	26.57%
REGYR 1978-80	N	74	0	4	4	82
	ROW%	90.24%	0.00%	4.88%	4.88%	100.00%
	COL%	27.41%	0.00%	10.00%	18.18%	24.48%
REGYR 1981-83	N	68	1	10	4	83
	ROW%	81.93%	1.20%	12.05%	4.82%	100.00%
	COL%	25.19%	33.33%	25.00%	18.18%	24.78%
REGYR 1984-86	N	58	0	20	3	81
	ROW%	71.60%	0.00%	24.69%	3.70%	100.00%
	COL%	21.48%	0.00%	50.00%	13.64%	24.18%
TOTAL	N	270	3	40	22	335
	ROW%	80.60%	0.90%	11.94%	6.57%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A6

Proportions of Men and Women for 1975-1986 Entrants:
Chemistry

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	72	16	1	89
	ROW%	80.90%	17.98%	1.12%	100.00%
	COL%	28.80%	19.05%	100.00%	26.57%
REGYR 1978-80	N	58	24	0	82
	ROW%	70.73%	29.27%	0.00%	100.00%
	COL%	23.20%	28.57%	0.00%	24.48%
REGYR 1981-83	N	61	22	0	83
	ROW%	73.49%	26.51%	0.00%	100.00%
	COL%	24.40%	26.19%	0.00%	24.78%
REGYR 1984-86	N	59	22	0	81
	ROW%	72.84%	27.16%	0.00%	100.00%
	COL%	23.60%	26.19%	0.00%	24.18%
TOTAL	N	250	84	1	335
	ROW%	74.63%	25.07%	0.30%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%

Table A7

Ethnic Composition for 1975-1986 Entrants[a]:
Mathematics

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	18	0	7	3	28
	ROW%	64.29%	0.00%	25.00%	10.71%	100.00%
	COL%	24.32%	0.00%	30.43%	75.00%	27.45%
REGYR 1978-80	N	21	0	1	1	23
	ROW%	91.30%	0.00%	4.35%	4.35%	100.00%
	COL%	28.38%	0.00%	4.35%	25.00%	22.55%
REGYR 1981-83	N	15	1	4	0	20
	ROW%	75.00%	5.00%	20.00%	0.00%	100.00%
	COL%	20.27%	100.00%	17.39%	0.00%	19.61%
REGYR 1984-86	N	20	0	11	0	31
	ROW%	64.52%	0.00%	35.48%	0.00%	100.00%
	COL%	27.03%	0.00%	47.83%	0.00%	30.39%
TOTAL	N	74	1	23	4	102
	ROW%	72.55%	0.98%	22.55%	3.92%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A8

Proportions of Men and Women for 1975-1986 Entrants:
Mathematics

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	22	6	0	28
	ROW%	78.57%	21.43%	0.00%	100.00%
	COL%	26.83%	30.00%	0.00%	27.45%
REGYR 1978-80	N	20	3	0	23
	ROW%	86.96%	13.04%	0.00%	100.00%
	COL%	24.39%	15.00%	0.00%	22.55%
REGYR 1981-83	N	17	3	0	20
	ROW%	85.00%	15.00%	0.00%	100.00%
	COL%	20.73%	15.00%	0.00%	19.61%
REGYR 1984-86	N	23	8	0	31
	ROW%	74.19%	25.81%	0.00%	100.00%
	COL%	28.05%	40.00%	0.00%	30.39%
TOTAL	N	82	20	0	102
	ROW%	80.39%	19.61%	0.00%	100.00%
	COL%	100.00%	100.00%	0.00%	100.00%

Table A9

Ethnic Composition for 1975-1986 Entrants[a]:
Physics

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	18	0	11	9	38
	ROW%	47.37%	0.00%	28.95%	23.68%	100.00%
	COL%	25.35%	0.00%	15.94%	81.82%	25.00%
REGYR 1978-80	N	19	1	16	0	36
	ROW%	52.78%	2.78%	44.44%	0.00%	100.00%
	COL%	26.76%	100.00%	23.19%	0.00%	23.68%
REGYR 1981-83	N	18	0	15	0	33
	ROW%	54.55%	0.00%	45.45%	0.00%	100.00%
	COL%	25.35%	0.00%	21.74%	0.00%	21.71%
REGYR 1984-86	N	16	0	27	2	45
	ROW%	35.56%	0.00%	60.00%	4.44%	100.00%
	COL%	22.54%	0.00%	39.13%	18.18%	29.61%
TOTAL	N	71	1	69	11	152
	ROW%	46.71%	0.66%	45.39%	7.24%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A10

Proportions of Men and Women for 1975-1986 Entrants:
Physics

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	28	6	4	38
	ROW%	73.68%	15.79%	10.53%	100.00%
	COL%	22.22%	27.27%	100.00%	25.00%
REGYR 1978-80	N	32	4	0	36
	ROW%	88.89%	11.11%	0.00%	100.00%
	COL%	25.40%	18.18%	0.00%	23.68%
REGYR 1981-83	N	28	5	0	33
	ROW%	84.85%	15.15%	0.00%	100.00%
	COL%	22.22%	22.73%	0.00%	21.71%
REGYR 1984-86	N	38	7	0	45
	ROW%	84.44%	15.56%	0.00%	100.00%
	COL%	30.16%	31.82%	0.00%	29.61%
TOTAL	N	126	22	4	152
	ROW%	82.89%	14.47%	2.63%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%

Table A11

Ethnic Composition for 1975-1986 Entrants[a]:
Counseling Psychology

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	54	6	4	8	72
	ROW%	75.00%	8.33%	5.56%	11.11%	100.00%
	COL%	40.30%	50.00%	80.00%	100.00%	45.28%
REGYR 1978-80	N	44	3	1	0	48
	ROW%	91.67%	6.25%	2.08%	0.00%	100.00%
	COL%	32.84%	25.00%	20.00%	0.00%	30.19%
REGYR 1981-83	N	21	3	0	0	24
	ROW%	87.50%	12.50%	0.00%	0.00%	100.00%
	COL%	15.67%	25.00%	0.00%	0.00%	15.09%
REGYR 1984-86	N	15	0	0	0	15
	ROW%	100.00%	0.00%	0.00%	0.00%	100.00%
	COL%	11.19%	0.00%	0.00%	0.00%	9.43%
TOTAL	N	134	12	5	8	159
	ROW%	84.28%	7.55%	3.14%	5.03%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A12

Proportions of Men and Women for 1975-1986 Entrants:
Counseling Psychology

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	23	49	0	72
	ROW%	31.94%	68.06%	0.00%	100.00%
	COL%	46.94%	44.55%	0.00%	45.28%
REGYR 1978-80	N	14	34	0	48
	ROW%	29.17%	70.83%	0.00%	100.00%
	COL%	28.57%	30.91%	0.00%	30.19%
REGYR 1981-83	N	10	14	0	24
	ROW%	41.67%	58.33%	0.00%	100.00%
	COL%	20.41%	12.73%	0.00%	15.09%
REGYR 1984-86	N	2	13	0	15
	ROW%	13.33%	86.67%	0.00%	100.00%
	COL%	4.08%	11.82%	0.00%	9.43%
TOTAL	N	49	110	0	159
	ROW%	30.82%	69.18%	0.00%	100.00%
	COL%	100.00%	100.00%	0.00%	100.00%

Table A13

Ethnic Composition for 1975-1986 Entrants[a]:
Clinical Psychology

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	20	0	1	13	34
	ROW%	58.82%	0.00%	2.94%	38.24%	100.00%
	COL%	17.39%	0.00%	50.00%	48.15%	23.29%
REGYR 1978-80	N	21	1	0	8	30
	ROW%	70.00%	3.33%	0.00%	26.67%	100.00%
	COL%	18.26%	50.00%	0.00%	29.63%	20.55%
REGYR 1981-83	N	39	0	1	3	43
	ROW%	90.70%	0.00%	2.33%	6.98%	100.00%
	COL%	33.91%	0.00%	50.00%	11.11%	29.45%
REGYR 1984-86	N	35	1	0	3	39
	ROW%	89.74%	2.56%	0.00%	7.69%	100.00%
	COL%	30.43%	50.00%	0.00%	11.11%	26.71%
TOTAL	N	115	2	2	27	146
	ROW%	78.77%	1.37%	1.37%	18.49%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A14

Proportions of Men and Women for 1975-1986 Entrants:
Clinical Psychology

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	14	20	0	34
	ROW%	41.18%	58.82%	0.00%	100.00%
	COL%	24.14%	22.99%	0.00%	23.29%
REGYR 1978-80	N	15	14	1	30
	ROW%	50.00%	46.67%	3.33%	100.00%
	COL%	25.86%	16.09%	100.00%	20.55%
REGYR 1981-83	N	20	23	0	43
	ROW%	46.51%	53.49%	0.00%	100.00%
	COL%	34.48%	26.44%	0.00%	29.45%
REGYR 1984-86	N	9	30	0	39
	ROW%	23.08%	76.92%	0.00%	100.00%
	COL%	15.52%	34.48%	0.00%	26.71%
TOTAL	N	58	87	1	146
	ROW%	39.73%	59.59%	0.68%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%

Table A15

Ethnic Composition for 1975-1986 Entrants[a]:
Sociology

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	26	5	4	8	43
	ROW%	60.47%	11.63%	9.30%	18.60%	100.00%
	COL%	30.59%	41.67%	22.22%	57.14%	33.33%
REGYR 1978-80	N	22	3	5	3	33
	ROW%	66.67%	9.09%	15.15%	9.09%	100.00%
	COL%	25.63%	25.00%	27.78%	21.43%	25.58%
REGYR 1981-83	N	15	2	4	1	22
	ROW%	68.18%	9.09%	18.18%	4.55%	100.00%
	COL%	17.65%	16.67%	22.22%	7.14%	17.05%
REGYR 1984-86	N	22	2	5	2	31
	ROW%	70.97%	6.45%	16.13%	6.45%	100.00%
	COL%	25.88%	16.67%	27.78%	14.29%	24.03%
TOTAL	N	85	12	18	14	129
	ROW%	65.89%	9.30%	13.95%	10.85%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A16

Proportions of Men and Women for 1975-1986 Entrants:
Sociology

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	19	23	1	43
	ROW%	44.19%	53.49%	2.33%	100.00%
	COL%	35.85%	30.67%	100.00%	33.33%
REGYR 1978-80	N	14	19	0	33
	ROW%	42.42%	57.58%	0.00%	100.00%
	COL%	26.42%	25.33%	0.00%	25.58%
REGYR 1981-83	N	9	13	0	22
	ROW%	40.91%	59.09%	0.00%	100.00%
	COL%	16.98%	17.33%	0.00%	17.05%
REGYR 1984-86	N	11	20	0	31
	ROW%	35.48%	64.52%	0.00%	100.00%
	COL%	20.75%	26.67%	0.00%	24.03%
TOTAL	N	53	75	1	129
	ROW%	41.09%	58.14%	0.78%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%

Table A17

Ethnic Composition for 1975-1986 Entrants[a]:
Theatre

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	18	1	1	7	27
	ROW%	66.67%	3.70%	3.70%	25.93%	100.00%
	COL%	40.00%	100.00%	25.00%	87.50%	46.55%
REGYR 1978-80	N	14	0	0	1	15
	ROW%	93.33%	0.00%	0.00%	6.67%	100.00%
	COL%	31.11%	0.00%	0.00%	12.50%	25.86%
REGYR 1981-83	N	10	0	1	0	11
	ROW%	90.91%	0.00%	9.09%	0.00%	100.00%
	COL%	22.22%	0.00%	25.00%	0.00%	18.97%
REGYR 1984-86	N	3	0	2	0	5
	ROW%	60.00%	0.00%	40.00%	0.00%	100.00%
	COL%	6.67%	0.00%	50.00%	0.00%	8.62%
TOTAL	N	45	1	4	8	58
	ROW%	77.59%	1.72%	6.90%	13.79%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A18

Proportions of Men and Women for 1975-1986 Entrants:
Theatre

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	10	17	0	27
	ROW%	37.04%	62.96%	0.00%	100.00%
	COL%	35.71%	56.67%	0.00%	46.55%
REGYR 1978-80	N	7	8	0	15
	ROW%	46.67%	53.33%	0.00%	100.00%
	COL%	25.00%	26.67%	0.00%	25.86%
REGYR 1981-83	N	8	3	0	11
	ROW%	72.73%	27.27%	0.00%	100.00%
	COL%	28.57%	10.00%	0.00%	18.97%
REGYR 1984-86	N	3	2	0	5
	ROW%	60.00%	40.00%	0.00%	100.00%
	COL%	10.71%	6.67%	0.00%	8.62%
TOTAL	N	28	30	0	58
	ROW%	48.28%	51.72%	0.00%	100.00%
	COL%	100.00%	100.00%	0.00%	100.00%

Table A19

Ethnic Composition for 1975-1986 Entrants[a]:
English

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	29	1	1	6	37
	ROW%	78.38%	2.70%	2.70%	16.22%	100.00%
	COL%	32.58%	33.33%	11.11%	75.00%	33.94%
REGYR 1978-80	N	21	2	3	0	26
	ROW%	80.77%	7.69%	11.54%	0.00%	100.00%
	COL%	23.60%	66.67%	33.33%	0.00%	23.85%
REGYR 1981-83	N	10	0	2	1	13
	ROW%	76.92%	0.00%	15.38%	7.69%	100.00%
	COL%	11.24%	0.00%	22.22%	12.50%	11.93%
REGYR 1984-86	N	29	0	3	1	33
	ROW%	87.88%	0.00%	9.09%	3.03%	100.00%
	COL%	32.58%	0.00%	33.33%	12.50%	30.23%
TOTAL	N	89	3	9	8	109
	ROW%	81.65%	2.75%	8.26%	7.34%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A20

Proportions of Men and Women for 1975-1986 Entrants:
English

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	14	23	0	37
	ROW%	37.84%	62.16%	0.00%	100.00%
	COL%	29.79%	37.10%	0.00%	33.94%
REGYR 1978-80	N	14	12	0	26
	ROW%	53.85%	46.15%	0.00%	100.00%
	COL%	29.79%	19.35%	0.00%	23.85%
REGYR 1981-83	N	5	8	0	13
	ROW%	38.46%	61.54%	0.00%	100.00%
	COL%	10.64%	12.90%	0.00%	11.93%
REGYR 1984-86	N	14	19	0	33
	ROW%	42.42%	57.58%	0.00%	100.00%
	COL%	29.79%	30.65%	0.00%	30.28%
TOTAL	N	47	62	0	109
	ROW%	43.12%	56.88%	0.00%	100.00%
	COL%	100.00%	100.00%	0.00%	100.00%

83

80

Table A21
Ethnic Composition for 1975-1986 Entrants[a]:
History

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	22	0	3	12	37
	ROW%	59.46%	0.00%	8.11%	32.43%	100.00%
	COL%	23.16%	0.00%	42.86%	85.71%	29.60%
REGYR 1978-80	N	18	5	2	0	25
	ROW%	72.00%	20.00%	8.00%	0.00%	100.00%
	COL%	18.95%	55.56%	28.57%	0.00%	20.00%
REGYR 1981-83	N	23	3	1	1	28
	ROW%	82.14%	10.71%	3.57%	3.57%	100.00%
	COL%	24.21%	33.33%	14.29%	7.14%	22.40%
REGYR 1984-86	N	32	1	1	1	35
	ROW%	91.43%	2.86%	2.86%	2.86%	100.00%
	COL%	33.68%	11.11%	14.29%	7.14%	28.00%
TOTAL	N	95	9	7	14	125
	ROW%	76.00%	7.20%	5.60%	11.20%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A22

Proportions of Men and Women for 1975-1986 Entrants:
History

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	19	18	0	37
	ROW%	51.35%	48.65%	0.00%	100.00%
	COL%	28.36%	31.03%	0.00%	29.60%
REGYR 1978-80	N	12	13	0	25
	ROW%	48.00%	52.00%	0.00%	100.00%
	COL%	17.91%	22.41%	0.00%	20.00%
REGYR 1981-83	N	17	11	0	28
	ROW%	60.71%	39.29%	0.00%	100.00%
	COL%	25.37%	18.97%	0.00%	22.40%
REGYR 1984-86	N	19	16	0	35
	ROW%	54.29%	45.71%	0.00%	100.00%
	COL%	28.36%	27.59%	0.00%	28.00%
TOTAL	N	67	58	0	125
	ROW%	53.60%	46.40%	0.00%	100.00%
	COL%	100.00%	100.00%	0.00%	100.00%

Table A23

Ethnic Composition for 1975-1986 Entrants[a]:
Political Science

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	26	4	10	11	51
	ROW%	50.98%	7.84%	19.61%	21.57%	100.00%
	COL%	32.50%	30.77%	22.22%	91.67%	34.00%
REGYR 1978-80	N	23	4	13	0	40
	ROW%	57.50%	10.00%	32.50%	0.00%	100.00%
	COL%	28.75%	30.77%	28.89%	0.00%	26.67%
REGYR 1981-83	N	12	4	9	1	26
	ROW%	46.15%	15.38%	34.62%	3.85%	100.00%
	COL%	15.00%	30.77%	20.00%	8.33%	17.33%
REGYR 1984-86	N	19	1	13	0	33
	ROW%	57.58%	3.03%	39.39%	0.00%	100.00%
	COL%	23.75%	7.69%	28.89%	0.00%	22.00%
TOTAL	N	80	13	45	12	150
	ROW%	53.33%	8.67%	30.00%	8.00%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A24

Proportions of Men and Women for 1975-1986 Entrants:
Political Science

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	35	16	0	51
	ROW%	68.63%	31.37%	0.00%	100.00%
	COL%	32.41%	38.10%	0.00%	34.00%
REGYR 1978-80	N	31	9	0	40
	ROW%	77.50%	22.50%	0.00%	100.00%
	COL%	28.70%	21.43%	0.00%	26.67%
REGYR 1981-83	N	17	9	0	26
	ROW%	65.38%	34.62%	0.00%	100.00%
	COL%	15.74%	21.43%	0.00%	17.33%
REGYR 1984-86	N	25	8	0	33
	ROW%	75.76%	24.24%	0.00%	100.00%
	COL%	23.15%	19.05%	0.00%	22.00%
TOTAL	N	108	42	0	150
	ROW%	72.00%	28.00%	0.00%	100.00%
	COL%	100.00%	100.00%	0.00%	100.00%

Table A25

Ethnic Composition for 1975-1986 Entrants[a]:
Economics

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	33	3	16	13	65
	ROW%	50.77%	4.62%	24.62%	20.00%	100.00%
	COL%	19.64%	60.00%	20.00%	56.52%	23.55%
REGYR 1978-80	N	62	0	19	4	85
	ROW%	72.94%	0.00%	22.35%	4.71%	100.00%
	COL%	36.90%	0.00%	23.75%	17.39%	30.80%
REGYR 1981-83	N	39	1	24	2	66
	ROW%	59.09%	1.52%	36.36%	3.03%	100.00%
	COL%	23.21%	20.00%	30.00%	8.70%	23.91%
REGYR 1984-86	N	34	1	21	4	60
	ROW%	56.67%	1.67%	35.00%	6.67%	100.00%
	COL%	20.24%	20.00%	26.25%	17.39%	21.74%
TOTAL	N	168	5	80	23	276
	ROW%	60.87%	1.81%	28.99%	8.33%	100.00%
	COL%	100.00%	100.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A26

Proportions of Men and Women for 1975-1986 Entrants:
Economics

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	57	8	0	65
	ROW%	87.69%	12.31%	0.00%	100.00%
	COL%	26.27%	13.56%	0.00%	23.55%
REGYR 1978-80	N	61	24	0	85
	ROW%	71.76%	28.24%	0.00%	100.00%
	COL%	28.11%	40.68%	0.00%	30.80%
REGYR 1981-83	N	54	12	0	66
	ROW%	81.82%	18.18%	0.00%	100.00%
	COL%	24.88%	20.34%	0.00%	23.91%
REGYR 1984-86	N	45	15	0	60
	ROW%	75.00%	25.00%	0.00%	100.00%
	COL%	20.74%	25.42%	0.00%	21.74%
TOTAL	N	217	59	0	276
	ROW%	78.62%	21.38%	0.00%	100.00%
	COL%	100.00%	100.00%	0.00%	100.00%

Table A27

Ethnic Composition for 1975-1986 Entrants[a]:
Philosophy

		WHITE	BLACK	FOREIGN	OTHER & MISSING	TOTAL
REGYR 1975-77	N	10	0	1	5	16
	ROW%	62.50%	0.00%	6.25%	31.25%	100.00%
	COL%	16.13%	0.00%	25.00%	83.33%	22.22%
REGYR 1978-80	N	26	0	0	0	26
	ROW%	100.00%	0.00%	0.00%	0.00%	100.00%
	COL%	41.94%	0.00%	0.00%	0.00%	36.11%
REGYR 1981-83	N	9	0	2	1	12
	ROW%	75.00%	0.00%	16.67%	8.33%	100.00%
	COL%	14.52%	0.00%	50.00%	16.67%	16.67%
REGYR 1984-86	N	17	0	1	0	18
	ROW%	94.44%	0.00%	5.56%	0.00%	100.00%
	COL%	27.42%	0.00%	25.00%	0.00%	25.00%
TOTAL	N	62	0	4	6	72
	ROW%	86.11%	0.00%	5.56%	8.33%	100.00%
	COL%	100.00%	0.00%	100.00%	100.00%	100.00%

[a]"Other and missing" includes Hispanics, Asians, and Native Americans. Prior to 1976 most students without ethnic codes are probably White (see text).

Table A28

Proportions of Men and Women for 1975-1986 Entrants:
Philosophy

		MALE	FEMALE	MISSING GENDER	TOTAL
REGYR 1975-77	N	13	3	0	16
	ROW%	81.25%	18.75%	0.00%	100.00%
	COL%	21.31%	27.27%	0.00%	22.22%
REGYR 1978-80	N	22	4	0	26
	ROW%	84.62%	15.38%	0.00%	100.00%
	COL%	36.07%	36.36%	0.00%	36.11%
REGYR 1981-83	N	10	2	0	12
	ROW%	83.33%	16.67%	0.00%	100.00%
	COL%	16.39%	18.18%	0.00%	16.67%
REGYR 1984-86	N	16	2	0	18
	ROW%	88.89%	11.11%	0.00%	100.00%
	COL%	26.23%	18.18%	0.00%	25.00%
TOTAL	N	61	11	0	72
	ROW%	84.72%	15.28%	0.00%	100.00%
	COL%	100.00%	100.00%	0.00%	100.00%

Appendix B - Survival Analysis

To develop the survival analysis model used here, we start out by assuming a piecewise exponential distribution of survival times within each graduate program. This implies, for each of K programs, a constant hazard, θ_{ik} ($k = 1, 2, \ldots K$; $i = 1, 2 \ldots I$), within each of I one-year time intervals. Let d_{ikj} be an indicator variable such that $d_{ikj} = 1$ if person j in program k experiences the event (candidacy or graduation) in the i^{th} interval; otherwise, $d_{ikj} = 0$. Let t_{ikj} be the amount of time person j in program k spends in the i^{th} interval. Let $d_{ik} = \sum_{j=1}^{J} d_{ikj}$ be the number of events that occur in interval i for program k and let $t_{ik} = \sum_{j=1}^{J} t_{ikj}$ be the total exposure time in interval i for members of graduate program k. We want to estimate the IK values of θ_{ik}. The likelihood for this model is

$$L(\underline{\theta}) = \prod_{k=1}^{K} \prod_{j=1}^{J} \prod_{i=1}^{I} \theta_{ik}^{d_{ikj}} \exp(-\theta_{ik} t_{ikj})$$

$$= \prod_{k=1}^{K} \prod_{i=1}^{I} \theta_{ik}^{d_{ik}} \exp(-\theta_{ik} t_{ik}) \qquad [1]$$

As demonstrated by Laird and Olivier (1981, p. 235) in the case of a simpler model, the likelihood obtained by assuming separate piecewise exponential distributions within programs is proportional to the likelihood that would be obtained under the assumption that each d_{ik} is an independent Poisson variate, conditional on t_{ik}, with $E(d_{ik} | t_{ik}) = t_{ik}\theta_{ik}$. That is,

$$L_p(\underset{\sim}{\theta}) = \prod_{i=1}^{I}\prod_{k=1}^{K} (t_{ik}\theta_{ik})^{d_{ik}} \exp(-t_{ik}\theta_{ik})/d_{ik}!$$

$$= \prod_{i=1}^{I}\prod_{k=1}^{K} t_{ik}^{d_{ik}}/d_{ik}! \prod_{i=1}^{I}\prod_{k=1}^{K} \theta_{ik}^{d_{ik}} \exp(-t_{ik}\theta_{ik})$$

$$\propto L(\underset{\sim}{\theta}) \qquad [2]$$

Because the likelihood kernels are the same, the two models can be used interchangeably for making likelihood-based inferences about the parameters θ_{ik}.

The maximum likelihood estimate of θ_{ik} is simply the occurrence rate for program k in interval i, d_{ik}/t_{ik}. In our analyses, we used a conventional life table approximation for the total exposure time for program k in interval i: $\tilde{t}_{ik} = n_{ik} - \frac{d_{ik} + c_{ik}}{2}$, where n_{ik} is the number of students in program k who had not yet experienced the event of interest as of the beginning of the i^{th} interval and c_{ik} is the number of students in program k who were censored during the i^{th} interval (see Laird & Oliver, p. 236).

A problem with ratios of occurrence to exposure, like $d_{ik}/\tilde{t}_{ik}$ or d_{ik}/t_{ik}, is that they tend to be unstable when sample sizes are small. We therefore wish to incorporate prior information about the parameters θ_{ik}. If we were to remain in the Poisson framework, the next step would be to assume a distribution conjugate to the Poisson for the θ_{ik}. Braun's (1985) approach, however, involves transforming the Poisson variates to normal variates and then applying empirical Bayes methods that have already been developed for the normal case. Let

$$X_{ik} = [d_{ik}/\tilde{t}_{ik}]^{1/2}$$

Then, if the Poisson assumption holds, we have approximately

$$\underset{\sim}{X}_k \sim N(\underset{\sim}{\mu}_k, \underset{\sim}{S}_k)$$

where $\underset{\sim}{X}_k = (X_{1k}, X_{2k}, \ldots X_{Ik})$, $\underset{\sim}{\mu}_k = (\theta_{1k}^{1/2}, \theta_{2k}^{1/2}, \ldots \theta_{Ik}^{1/2})$ and $\underset{\sim}{S}_k$ is a diagonal matrix with the i^{th} diagonal element equal to $(4\tilde{\tau}_{ik})^{-1}$. The second level of the model assumes that the vectors $\underset{\sim}{\mu}_k$ are independently generated from a multivariate normal distribution, i.e.,

$$\underset{\sim}{\mu}_k \sim N(\underset{\sim}{\mu}^*, \underset{\sim}{\Sigma}^*), \quad k = 1, 2, \ldots K \quad .$$

We assume $\underset{\sim}{\mu}^*$ and $\underset{\sim}{\Sigma}^*$ are unknown and must be estimated from the data.

This model is a special case of the general regression model described in Braun, Jones, Rubin, and Thayer (1983). Braun et al. show how the EM algorithm (Dempster, Laird, & Rubin, 1977) can be used to obtain maximum likelihood estimates of $\underset{\sim}{\mu}_k^*$ and $\underset{\sim}{\Sigma}^*$ as well as the posterior distributions of the $\{\underset{\sim}{\mu}_k\}$ given these estimates and the data. The means of these posterior distributions provide estimates of the $\{\underset{\sim}{\mu}_k\}$. Squaring these estimates in turn yields estimates of the $\{\underset{\sim}{\theta}_k\}$.

The estimation procedure for the EB survival analysis developed by Braun (1985) differs in two ways from the general regression model of Braun et al. (1983). First, the values of $\mathrm{Var}(X_{ik})$ are known in the present case and need not be re-estimated in the M step of the EM algorithm. Second, to reduce the number of parameters to be estimated, a special structure is assumed for $\underset{\sim}{\Sigma}^*$, as follows:

$$\underset{\sim}{\Sigma}^* = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots \\ \rho & 1 & \rho & \cdots \\ \rho^2 & \rho & 1 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix} \quad . \qquad [3]$$

That is, the correlations between the square roots of the hazards are assumed to be geometrically decreasing. This requires that the computational algorithm be modified to obtain maximum likelihood estimates of σ^2 and ρ (Szatrowski, 1976). Based on a preliminary investigation of the robustness of the estimation procedure to the assumption of the covariance structure in Equation 3, Braun (1985) concluded that the obtained estimates would not be expected to vary greatly over a reasonable collection of assumed covariance structures. In the current study, the obtained estimates of ρ were .61 for the graduation analysis, .31 for the candidacy analysis, and .60 for the analysis of graduation, given candidacy.

For the piecewise exponential survival model with intervals of length Δ_i, the probability of surviving through interval i_o for an individual in graduate program k is estimated by

$$\hat{S}_k(i_o) = \prod_{i \leq i_o} \exp(-\hat{\theta}_{ik}\Delta_i) \qquad [4]$$

This expression is equal to $\prod_{i \leq i_o} \exp(-\hat{\theta}_{ik})$ if $\Delta_i = 1$, $i = 1, 2 \ldots I$, as in the present case. The classical survival curves are obtained by setting $\hat{\theta}_{ik}$ equal to $d_{ik}/\tilde{t}_{ik}$; the EB curves are found by substituting the EB estimates of the hazards.

The differences between the classical and EB estimates were more apparent when sample sizes were small, as in survival analyses (not shown) that included only White students who were U.S. citizens. In these analyses, the classical hazard estimates showed wild fluctuations, whereas the EB estimates, which borrow strength from the remaining graduate programs, were smoother and better behaved. The EB survival functions were also smoother and

closer together than their classical counterparts. Of course, the more pleasing appearance of the EB graphs does not, in itself, demonstrate that these estimates are superior. However, Braun (1985) presents two types of evidence that support the superiority of the EB appproach. First, a cross-validation study of the methodology used here was conducted. Data were divided in half at random and the EB estimates of a set of survival curves, based on a half-sample, were compared with the classical estimates based on each of the two half-samples. Each of the EB curves nearly bisected the two more variable curves based on the classical approach, indicating that the EB method successfully borrowed information to provide more stable estimates. In a second analysis, Braun investigated the properties of a fully Bayes survival analysis method closely related to the present approach. Bayes and classical estimates of hazard functions for a truncated data set were compared to classical estimates based on the full data set. The Bayes estimates for the truncated data were found to reproduce more closely the classical estimates based on the full data than did the classical estimates based on the truncated data.

Appendix C

An Empirical Bayes Strategy for Logistic Regression

A simplified EB strategy for logistic regression, which, like Braun's (1985) survival analysis approach, takes advantage of existing EB methods for the normal case, is as follows: For each of the K graduate programs, obtain vectors of regression coefficients $\hat{\underline{B}}_k$ and their asymptotic covariance matrices $\underline{S}_k$ from ordinary maximum likelihood logistic regression. Make use of the fact that the $\hat{\underline{B}}_k$ are asymptotically normal and treat the $\underline{S}_k$ as known. Thus, we have

$$\hat{\underline{B}}_k \sim N(\underline{B}_k, \underline{S}_k).$$

Now assume a normal prior for the $\underline{B}_k$:

$$\underline{B}_k \sim N(\underline{\mu}^*, \underline{\Sigma}^*)$$

and get $E(\underline{B}_k | \hat{\underline{B}}_k, \hat{\underline{\mu}}^*, \hat{\underline{\Sigma}}^*)$, where $\hat{\mu}^*$ and $\hat{\Sigma}^*$ are MLES of μ^* and Σ^*, using the EM algorithm. This approach is very similar to that of Korn and Whittemore (1979). A more rigorous EB approach to logistic regression has been developed by Wong and Mason (1985).

297981

BEST COPY AVAILABLE